Behavioural Classification of Passengers in an Airport Terminal

MSc Thesis

by

Luuk Rozema

January 2017

Behavioural Classification of Passengers in an Airport Terminal

MSc Thesis

By

Lukas Marius Rozema

in partial fulfilment of the requirements for the degree of

Master of Science
in Civil Engineering – Transport and Planning

at the Delft University of Technology,
to be defended publicly on Wednesday 25 January, 2017 at 16:00.

Supervisor: **Prof. dr. ir. S.P. Hoogendoorn**
Civil Engineering, Department of Transport and Planning

Thesis committee: Dr. ir. W. Daamen
Civil Engineering, Department of Transport and Planning

Dr. ir. B.G.H. Gorte
Civil Engineering, Department of Geoscience and Remote Sensing

Ir. R.J.F. Grosmann
Netherlands Aerospace Centre

Some parts of this thesis are confidential and have not been included in this public version

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Contact: lmrozema@outlook.com

TU Delft

Dedicated
to the loving memory of my grandfather
Klaas Philips

Preface

In front of you is the culmination of my education at the Delft University of Technology. Having obtained my bachelor's degree in Civil Engineering through the years 2011-2014, I started the master's degree programme in 2014. This thesis is the final piece of my work that is intended to grant me the Master of Science degree in Civil Engineering, specialised in Transport & Planning. The assignment worked on in this thesis is a joint assignment of the Dutch Aerospace Centre (NLR), Schiphol Group, and Delft University of Technology, combined in the Innovative Mainport Alliance (SIM).

During the approximately ten months I worked on this thesis, I have been able to work at both the NLR and Schiphol Group and so have been able to gain insight into both organisations. Having always been interested in Schiphol airport, working on this thesis has given me the great opportunity to gain more knowledge of the airport during several behind-the-scenes tours – both related and unrelated to my thesis work. Several people have helped me during my work, either on a professional or personal level. I would like to express my gratitude to them here.

First of all, Winnie Daamen, who has been my daily supervisor at the university. Every other week, we met to discuss the progress. She has provided me with useful and very detailed feedback, which has been very helpful. In addition, I would like to thank graduation committee chair Serge Hoogendoorn for his feedback and enthusiasm. Lastly, I would like to thank Ben Gorte, external supervisor, for his feedback.

Furthermore, I would like to thank Ronald Grosmann – also a member of the graduation committee – and Michel van Eenige, who have supervised my work on behalf of the NLR. Special thanks go out to Astrid Dijkstra and Niels Bakker of Schiphol Group for introducing me to the world of Schiphol and connecting me to the right people within the organization.

Lastly, I want to thank my family, friends, and colleagues who have supported me along the way; your support has been indispensable. And of course, I want to thank you, the reader, for taking the time to read my thesis report. I hope you enjoy reading it. Should you have any remarks or questions, feel free to contact me; you can find my contact information in the front matter of this report.

Luuk Rozema
Amsterdam, January 2017

This page intentionally left blank.

Summary

The European PASSME project aims to provide passengers with a less stressful and more enjoyable experience at airports. Part of the project is the development of a new passenger demand forecast (PDF) system to accurately predict passenger demand on a short timescale. Contrary to current PDF systems, the PASSME PDF aims to integrate multiple sources of data. Data from a mobile app, sensor data, airport data, and airline data, enhanced with detailed information about the passenger process and airport process are combined to create a detailed PDF that is able to forecast on a scale as small as 20 to 30 minutes ahead, which is much smaller than current PDF systems. One part of the PASSME PDF is a behavioural simulation, for which a method for behavioural classification of passengers should be developed. Based on the mentioned data sources, passengers should be classified into a behavioural class, which can then be used in the behavioural simulation of the PDF.

This thesis concerns the creation of a passenger classification framework that is able to perform such behavioural classification. As such, the main research question has been defined as follows: "How can individual airport passengers be classified according to (visual) sensor-obtained personal characteristics in behavioural classes that can be used to predict passenger behaviour?". To answer this question, the subject has been broken down into three interacting blocks: Sensing, Processing, and Modelling. The Sensing block concerns acquiring data about passengers through sensors and other data sources. Next, in the Processing block, this data is processed into behavioural classes. Lastly, the behavioural classes are used in a behavioural model in the Modelling block. The main focus of the thesis is on the processing block. However, in order to address this block, the other two blocks should be discussed first.

Sensing can be broken down into three main subjects: the passenger process, passenger characteristics and their effect on behaviour, and collecting passenger data. The passenger process concerns the various interconnected processes a passenger has to go through at the airport. This includes processes such as check-in, passing security, and finally boarding the plane. Between the various mandatory parts of the process, passengers are free to perform discretionary activities in the terminal. The passenger process is different for departing, arriving, and transferring passengers. Hence, these three types of passengers should be treated separately during behavioural classification.

Passengers should be classified into behavioural classes based on their characteristics. As such, two types of characteristics are defined: passenger characteristics and behavioural characteristics. Behavioural characteristics are characteristics that are regarded as behaviour. Essentially, these are the characteristics that are to be predicted by the behavioural classification. The behavioural characteristics form the basis of behavioural classes. Passenger characteristics describe the passenger himself. Those characteristics are used to classify a passenger into a behavioural class. Both types of characteristics can be collected using various sources of data, such as camera data, radio frequency (RF) sensors, the airport database, and the airline database. Challenges associated with this are, however, combining these sources into one data set, containing information on an individual basis, and the privacy issues of a detailed data set on an individual level.

Regarding Modelling, passenger behaviour can be broken down into three levels of behaviour: strategic, tactical and operational. Activities are chosen on the strategic level. On the tactical levels, activities are (re-)ordered and the route to these activities is chosen. The operational level mainly pertains to walking behaviour, such as collision avoidance. Behavioural models most often model on one or two of these levels. Consequently, models greatly differ with respect to their model inputs. To effectively combine behavioural classification with a behavioural model, these inputs should be known.

The Processing block connects the Sensing and Modelling blocks; based on the collected passenger and behavioural characteristics, behavioural classes are formed. To perform this behavioural classification, the Clustering and Classification (CC) framework was developed, shown below in figure i.

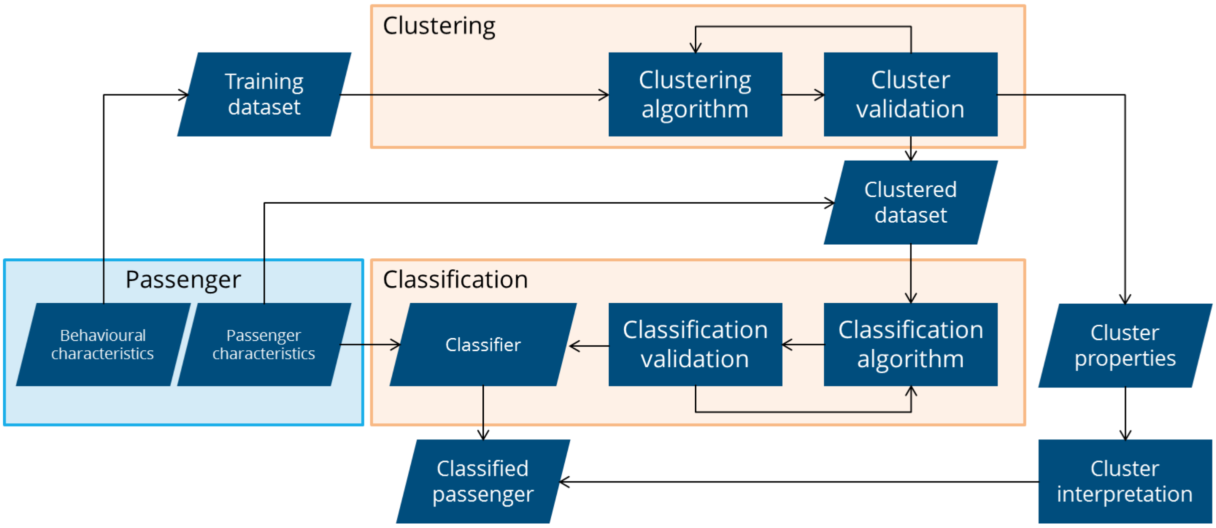


Figure i: The Clustering and Classification (CC) framework

The CC-framework consists of three main parts:

- **Passenger data:** The data input of the framework consists of the two types of characteristics: behavioural characteristics and passenger characteristics. This data is aggregated over a period of time in order to create a training data set that contains many passengers.
- **Clustering:** This part of the framework creates behavioural classes using Latent Class Analysis (LCA). The LCA creates behavioural classes solely based on the behavioural characteristics. After clustering, the training data also contains the class label for each passenger in the data set.
- **Classification:** Based on the passenger characteristics and the behavioural class to which each observation in the data set was assigned, a classifier is constructed using a classification algorithm. This classifier is then able to classify 'new' passengers into a behavioural class solely based on their passenger characteristics.

For both parts of the framework, several techniques are available. For the clustering part, Latent Class Analysis has been chosen, as this is a model-based technique that has proven to yield good results and that additionally has the benefit of providing good performance metrics that allow optimizing the behavioural classes formed by the method. The classification part of the algorithm uses the SAMME algorithm, which is a decision tree-based boosted ensemble classifier. Because decision trees are logic-based, the trees in the classifier can be easily interpreted. However, as the performance of decision trees is generally worse than some other algorithms, it was chosen to use boosting to increase accuracy. The CC-framework has been implemented using the R programming language.

To assess the performance of the framework, it has been applied to the PASSME data set. This data set contains several passenger and behavioural characteristics. The data set contains both departing and transferring passengers, totalling to about 4,000 observations, which is considered as an appropriate amount. In line with the aforementioned observation of differing passenger processes, the subsets of departing and transferring passengers have been treated as separate data sets. Analyses of this data set have been performed in order to assess the properties of the behavioural characteristics and possible relations between passenger characteristics and behavioural characteristics. Various relations were found, some of which are also confirmed by literature. Based on the data analysis, it was concluded that the PASSME is suitable for use with the CC-framework.

After analysis, the framework was applied to the PASSME data. Because the framework consists of two parts (clustering and classification), the results of clustering affect the results of classification. Both parts each have their own settings that affect the outcome. Therefore, a grid search has been performed in order to find the best classification performance. In this grid search, the number of classes, the bin size for discretisation of numerical behavioural attributes, and the maximum tree depth of the decision trees in the ensemble classifier were varied. The results for this grid search indicated that the classifier performance increases as the number of classes decreases and the bin size increases. The effect of the maximum tree depth is limited.

The grid search resulted in an optimum of two behavioural classes for both subsets. The resulting behavioural classes are very similar for both subsets and are primarily distinguished by the time that passengers spend at the gate, and the time between the decision of heading to the gate and flight departure. There are no significant differences between the other attributes.

Table i: Overview of behavioural classification results

Metric	Departing passengers	Transferring passengers
AUC	0.6883	0.6774
Overall accuracy	0.6363	0.6733
Average accuracy	0.6363	0.6733
Macro F1-score	0.6250	0.6138

The performance of the classification is shown in table i. The average accuracy for classification of departing passengers, indicating the average number of objects classified into the correct class, was found to be 64%, while the average accuracy of transferring passengers was found to be 67%. Although the classification meets the minimum requirement, set at an AUC of above 0.5, the classification performance cannot be regarded as very high. However, it should be kept in mind that the level of detail of the PASSME data was quite low and that the data set does not fully represent the data that would be available from actual (sensor) data. Consequently, it was concluded that the CC-framework yields promising results. Regarding possible further development of the framework, several recommendations have been made. Most importantly, it is recommended to use the framework with better data that is more alike to the actual data as it could be expected in combination with the PDF.

This page intentionally left blank.

Table of Contents

This page intentionally left blank.

List of tables

This page intentionally left blank.

List of figures

This page intentionally left blank.

Glossary

AAS	Amsterdam Airport Schiphol
Attribute	Feature of an object in a data set
Behavioural characteristic	Characteristic regarding the behaviour of a passenger; these characteristics are to be predicted by the behavioural classification
Classifier	Trained process that assigns an input object to a class
Cluster	Group of objects wherein each object is more similar to other objects within the cluster, than it is to objects not in the cluster
Discretionary activity	Non-mandatory activity, such as visiting a restaurant or inquiring information
Ensemble method	Combination of multiple (simple) classifiers to form a more accurate classifier
Object	An observation/tuple in a data set
Passenger characteristic	Characteristic describing the passenger; based on these characteristics, passengers are assigned to a behavioural class
PASSME	Personalised Airport Systems for Seamless Mobility & Experience, a European project aimed at improving the passenger experience at airports
PDF	Passenger demand forecast
PDT	Position determination technologies (PDT) is a general term for technologies that can provide the location of an object or person
SSBPC	Self-service boarding pass check, an automatic entry control system that requires passengers to self-scan their boarding pass

This page intentionally left blank.

1 Introduction

In the top 20 of busiest airports in the world, Amsterdam Airport Schiphol (AAS) ranks 14th with more than 58 million enplaning and deplaning passengers per year. Schiphol ranks 5th when only accounting for international passenger traffic , according to the ACI world rankings (Airports Council International (ACI), 4 April 2016). Recent news headlines such as “Extra staff Schiphol against long queues” (AT5, 22 April 2016) and “Schiphol advises travellers to arrive on time during the May-holidays” (NU.nl, 20 April 2016) highlight the fact that Schiphol is a busy airport that has to effectively deal with crowd management in order to serve such a high number of passengers safely, efficiently and satisfactorily. This means that adequate information is needed to be able to adjust operations to these high demands. To this end, Passenger Demand Forecast (PDF) systems are in place. These PDF-systems provide forecasts about the passenger demand at various areas in the terminal. However, the static nature of the data used in current systems limits the accuracy of these forecasts to a level which is not usable for adjusting and optimising operations during the day of operation itself.

In June 2015, the PASSME project was initiated. This European project pursues four main objectives with respect to improving the passenger experience in airports. One of these objectives is to “reduce door-to-door airport travel time for passengers in Europe by 60 minutes” (PASSME, n.d.-a). For this goal, a new proactive and real-time PDF-system is being developed. This new system aims to combine data from sensor observations, passenger behaviour data and data from the airport operations plan into a detailed forecast of the demand at various areas of the airport, such as the security filters or check-in desks (PASSME, n.d.-b). Whereas current PDF-systems accurately predict for a scope of a few days to a few hours in advance, the PASSME PDF-system will be also be able to forecast the coming twenty to thirty minutes. Passenger-centric operations, which have been planned on a quarterly, monthly and weekly basis, can be planned, controlled and optimised on the day of operation with this new information. Due to the optimal staffing that can be realised, passengers experience minimal delays during any of the passenger-centric operations. This will lead to an almost seamless experience for passengers in the terminal.

The new PDF-system will rely on multiple data sources and techniques to create an accurate forecast. One of the techniques that will be used is the behavioural classification of passengers. These behavioural classes represent classes of passengers having certain characteristic behaviour with respect to the choices they make and the walking behaviour they show. Based on sensor-obtained passenger characteristics, passengers will be classified into a specific behavioural class. This technique is yet to be developed. This thesis researches the possibilities of finding and defining behavioural classes of passengers and classifying individual passengers to such a behavioural class. Amsterdam Airport Schiphol is the main case for this thesis. Main questions that are involved in this subject are what kind of passenger characteristics can be obtained from sensor data, how clusters can be formed and verified, subsequently how passengers can be assigned to a behavioural class and what the corresponding accuracy is.

The subsequent sections in this chapter further outline the problem. First, the problem statement is presented in section 1.1. This includes the main research question, the objective of the thesis study, and the sub research questions. Section 1.2 details the scope of the research. Sections 1.3 and 1.4 specify the research contribution and methodology. Section 1.5 closes off the first chapter by giving an overview of the structure of the rest of the report.

1.1 Problem statement

This section further introduces the topic of this thesis report. First, the main research question is introduced. To support this research question, the objective of the research is then defined. Next, to clarify the various aspects of the research, the thesis is divided into three blocks of interest. Additional research sub-questions are defined based on these blocks.

1.1.1 Main research question

The main topic for this thesis is to categorise passenger into behavioural classes based on their characteristics and relate these to their behaviour. This results in the following main research question:

How can individual airport passengers be classified according to (visual) sensor-obtained personal characteristics in behavioural classes that can be used to predict passenger behaviour?

1.1.2 Objective

The objective of the thesis will be to create a passenger classification system. This passenger classification system will classify individual passengers in the airport terminal into a behavioural class, with a focus on using (visual) sensor-observed characteristics for classification. The aim of the behavioural classes is to predict the passenger's behaviour with respect to activity choice, route choice and walking behaviour.

The goal of the thesis work is threefold. The first goal is to develop a methodology to distinguish behavioural classes based on data regarding passenger behaviour and characteristics. The second goal is to develop a methodology to classify observed passengers to a behavioural class. The third and last goal is to create and implement a framework that can perform both aforementioned tasks in an integrated manner.

1.1.3 Research questions

So far, several aspects of interest for the thesis work have been mentioned. To clarify this, the thesis has been broken down into three interacting main blocks in figure 1.1: Sensing, Processing, and Modelling. These three blocks form the main areas of interest for the thesis, and based on these, further research questions have been formulated. The double-headed arrows between the blocks in the figure signify the relations between the different blocks; i.e. the blocks interact with respect to requirements, data, et cetera.

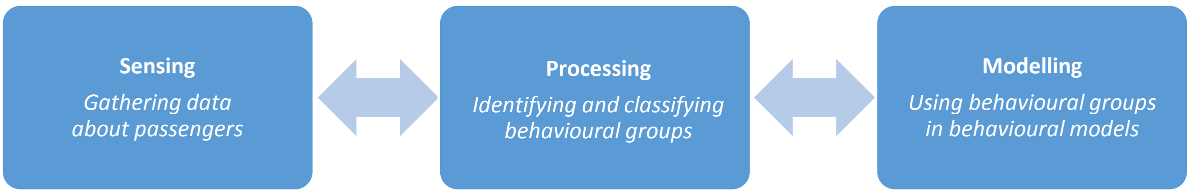


Figure 1.1: The three thesis blocks

1.1.3.1 Sensing
The first block is the Sensing block. This block pertains to the aspect of gathering information about passengers. This could be either via visual or non-visual sensors, airport information systems, or airline information systems. Additionally, this block focusses on the process that a passenger goes through at the airport. The sub-question related to this block has hence been formulated as follows.

- **Sensing sub-question:** Which passenger characteristics that can be used for behavioural classification can be obtained from sensors and information systems in an airport terminal?

1.1.3.2 Modelling
The last block is the Modelling block. This block pertains to the behavioural models and how these could interface with a behavioural classification as proposed in the thesis. However, because this aspect of the PASSME project has not been fully developed as of yet, this part of the thesis is more on a global, general level. The sub-question associated with this block is as follows.

- **Modelling sub-question:** How could behavioural classes form the input to behavioural models such as in a passenger demand forecast system?

1.1.3.3 Processing
The middle block, which is also the most important block, is the Processing block. The importance of this block follows from the position in the figure; it connects the Sensing and Modelling block. Consequently, this block forms the bridge between raw data from observations and the behavioural models in a passenger demand forecast system. The main subject of this block is the development of a framework to transform the passenger observations into behavioural classes and the subsequent generation of classification rules. As the processing block is the main focus of the thesis, multiple sub-questions have been formulated for this block:

- **Processing sub-question 1:** Which behavioural classes can be used to predict passenger behaviour?
- **Processing sub-question 2:** How can passengers be classified to these behavioural classes?
- **Processing sub-question 3:** When is the result of classification satisfactory?

1.2 Scope of research
The main focus of this thesis is on the processing block of figure 1.1. This pertains primarily to the relation between passenger characteristics and their behaviour. The focus for this is mostly on passenger characteristics that can be derived from visual detection systems and other sources of data within the airport.

As the main focus is on segmenting and classifying passengers into behavioural groups, the two building blocks Sensing and Modelling are mainly discussed to provide information about what information is realistically available for passenger classification and what information should ideally be available at airports. Additionally, as there is currently no data available for classification (in an operational sense), it is discussed what kind of data can and should be collected. For instance, camera footage from cameras placed in the terminal, observation based on footage from security cameras or questionnaires. However, the main focus remains on passenger classification. Furthermore, it will be studied how this passenger classification can interface with a yet to be implemented behavioural model in a passenger demand forecast system. Hence, the two blocks Sensing and Modelling are mainly to provide information about how the behavioural classification fits in the whole operation.

1.3 Research contribution
The thesis work aims to deliver a contribution to scientific knowledge, but also keeps in mind the practical relevance of the work. After all, the forecasts resulting from behavioural classification are to be used in a practical setting of the PASSME project. The relevance and contribution of the thesis work are explained in the next two subsections.

1.3.1 Scientific contribution

Literature that is currently available mostly pertains to sensing and modelling. With respect to sensing, literature is available, for example on the possibilities of identifying events in camera footage and tagging, tracking and tracing (Ouyang & Wang, 2012). Additionally, with respect to modelling, ample literature is available. Again, these two aspects of the thesis work are mainly there to provide an overview.

The scientific contribution of this thesis can mainly be attributed to the Processing block, which is where the scientific knowledge gap is mainly present. Techniques for segmenting data into classes, as well as classifying objects into such classes are separately well-described fields of science. However, both operations are usually treated separately and not in an integrated manner. Additionally, there are works available that relate the characteristics of passengers to their behaviour using discrete choice models, such as in the works of Kalakou and Moura (2015); X. Liu, Usher, and Strawderman (2014). However, to the best of the author's knowledge, techniques to find groups of airport passengers based on their behaviour have not been applied yet. Consequently, this thesis aims to fill in the knowledge gap by developing an integrated framework for segmenting passengers into behavioural classes and classifying new observations to these classes. The approach is not entirely theoretical as the framework is tested on an actual set of survey data collected at an airport within the PASSME project.

1.3.2 Practical contribution

The behavioural classification that is to be done in the thesis work will be used to create behavioural classes that can be used in the behavioural modelling in PASSME. These behavioural classes will allow creating a heterogeneous population in the simulation of a PASSME PDF. This can help improving simulation accuracy as each detected passenger is assigned a specific set of behaviour that can recreate their behaviour more realistically.

1.4 Research methodology

The previous sections have presented the facets of the topic that will be researched. To satisfactorily answer the main question and sub-questions of the problem at hand, the research has been split into various phases. These phases of the research methodology are explained below.

1.4.1 Site visits and literature review

The thesis work has started off with site visits and conversations with involved individuals. These activities are mostly focussed on getting acquainted with the process that passengers go through at the airport. Additionally, literature has been reviewed with respect to Sensing and Modelling. This part of the research hence answers the research sub-questions associated with these blocks. Moreover, this part also concerns the processing block as possible techniques for performing behavioural classification have been researched. Based on the findings of this part, the requirements for the framework for behavioural classification that is to be developed were determined.

1.4.2 Development of a clustering methodology

To find groups, or clusters, in passenger behaviour, additional research in literature has been done to find techniques that could be appropriate for this goal. One technique has been selected, implemented, and tested on an actual data set.

1.4.3 Development of a classification methodology

For classification, a similar method has been performed. First, techniques have been researched using literature, keeping in mind that the technique should be 'compatible' with the clustering methodology. One technique was then chosen, implemented, and tested on the same data set, using the results from the clustering.

1.4.4 Integrating clustering and classification

During this part of the research, the clustering and classification techniques have been combined into one complete framework. This has been applied as a whole on the data set. Additionally, different parameters of the framework have been optimized by collecting the results of various ranges of parameter values and determining the best result. The implementation of clustering and classification answers the research question related to the processing block.

1.4.5 Formulating conclusions and recommendations

Based on the previous parts of the research, conclusions and recommendations for further improvements are made. Furthermore, limitations of the proposed framework are discussed.

1.5 Structure of the report

The subsequent chapters of this report aim to solve the problem and research questions that were introduced in this chapter. The structure of these chapters is as shown in figure 1.2. Each chapter of the report is visualized as a separate block with several interconnected topics. The light blue blocks indicate the main topic of the chapter. The orange strips on the right hand side of the figure indicate which of the thesis blocks are the most relevant for the chapter.

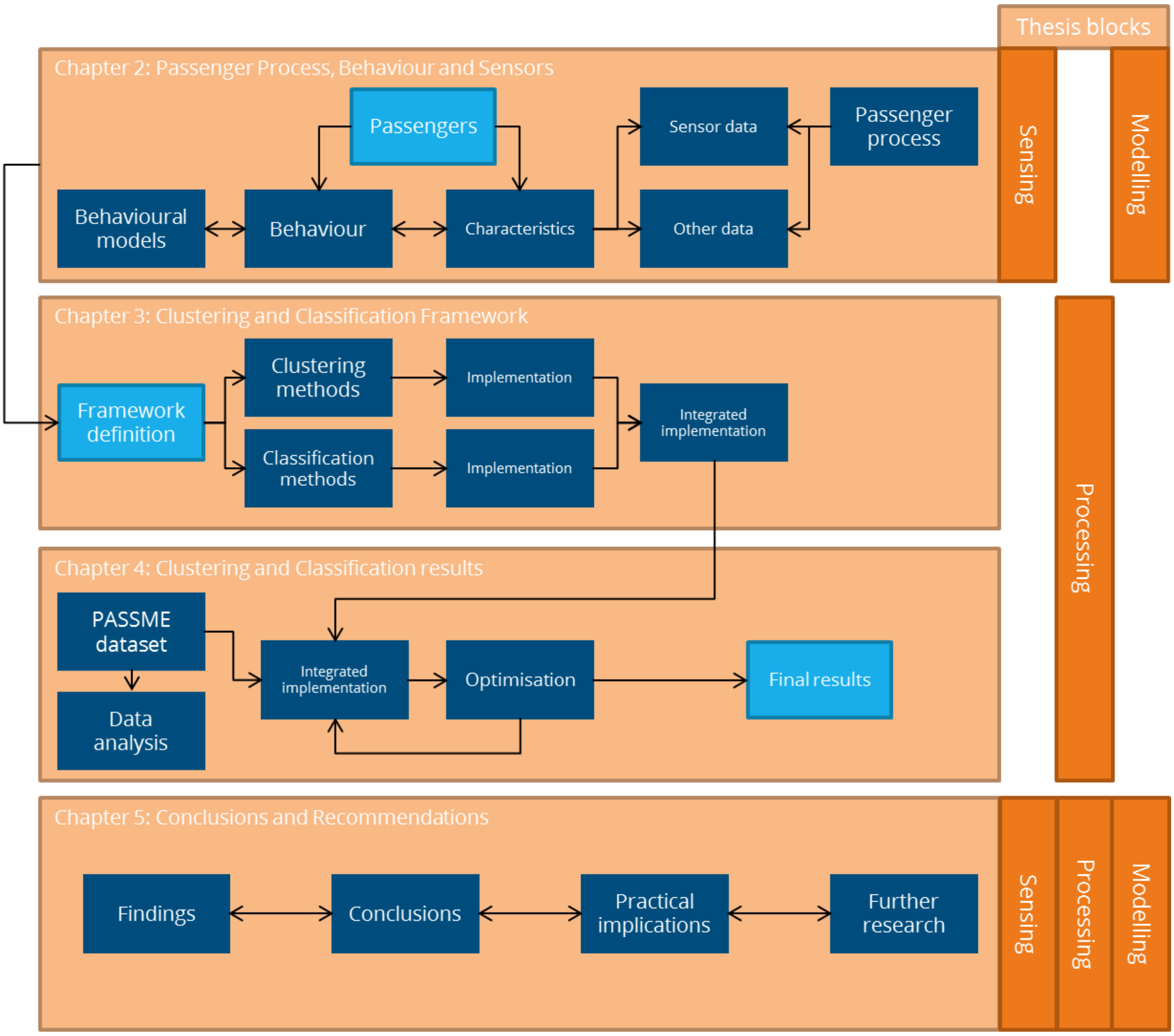


Figure 1.2: Structure of the report

The main topic of **chapter 2** is passengers. This chapter focusses on the sensing and modelling block of the thesis. For the sensing block, this chapter starts out with the passenger process that departing, transferring, and arriving passengers go through. Next, passenger characteristics and their relation to passenger behaviour are discussed based on literature. Sensor data sources and other data sources can provide information about passenger characteristics and their passenger process, which are discussed next. Lastly, the modelling block is discussed, introducing some principles of behavioural models and the levels of passenger behaviour.

Having addressed the sensing and modelling block in chapter 2, **chapter 3** shifts the focus to the processing block. The main topic in this chapter is the definition of a framework to find behavioural classes and perform classification, based on the types of characteristics set out in chapter 2. The framework definition is broken into two main parts: clustering and classification. Clustering divides passengers into behavioural classes, while classification creates the rules for assigning passengers to such classes. For both methods, various techniques and their characteristics are discussed, along with how to assess their performance. Finally both methods are implemented separately and later in an integrated manner. The specifics of this integrated application are then discussed.

Chapter 4 continues with the integrated framework of chapter 3 and still pertains to the processing block. The chapter introduces the PASSME data set, which contains passenger characteristics and behavioural characteristics. The data are analysed to find possible patterns and correlations in the data, based on which an expectation with respect to behavioural classes to be found can be made. Next, using the integrated implementation from chapter 3, the clustering and classification are applied to the PASSME data set. Because there are several parameters that can be adjusted, the results are optimised using a grid search. The optimal results are finally presented in the last part of the chapter.

Chapter 5, the last chapter of the report, takes into account the finding of all previous chapters and formulates the findings, conclusions, and recommendations for further research.

2

Passenger Process, Behaviour and Sensors

The previous chapter has introduced the subject of this thesis and the three conceptual blocks into which the topic has been divided. This chapter mostly pertains to the first and last blocks: Sensing and Modelling. It thereby aims to provide the background knowledge that is required to be able create a methodology for passenger classification.

The PASSME project, briefly introduced in the previous chapter, aims to improve the passenger experience. One of the project's goals is to develop a new PDF system, as laid out in the work of Grosmann (2016). In contrast with currently used PDF systems, the PASSME PDF aims to integrate multiple sources of data and use continuous, detailed monitoring of individual passenger's movement. Data about passengers in the terminal are collected and analysed, and a forecast is made. This concept is shown in figure 2.1.

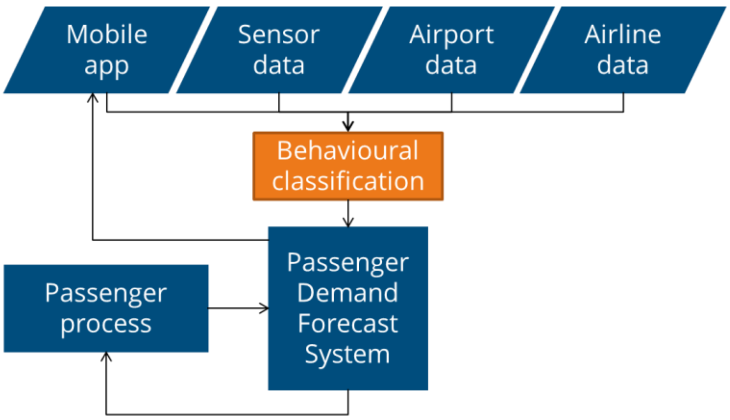


Figure 2.1: PASSME PDF and behavioural classification, adapted from Grosmann (2016)

As shown in the figure, the PDF uses multiple sources of data. An app on passengers' mobile devices can provide information about the passenger to the system. In return, the system can help personalise the passenger experience by providing personal advice to the passenger. Sensor data from visual sensors such as cameras, or systems such as Wi-Fi tracking, airport data, and airline data also provide information about the passenger. All this data is used for the PDF, which creates a passenger demand forecast in interaction with the passenger process, which describes the processes a passenger goes through. One of the aspects of the PDF is the modelling of passenger behaviour by using a behavioural model. This is where the behavioural classification of this thesis comes into play: based on the characteristics from the data fed into the PDF, a passenger should be assigned to a behavioural class that has a specific behaviour. These behavioural classes are then input to the behavioural model of the PDF.

This chapter further elaborates on the different aspects of the PDF as displayed in the figure. First, the passenger process is discussed in section 2.1. This passenger process encompasses all the (mandatory) steps that passengers have to go through from arriving at the terminal until the moment of boarding the aircraft. After that, the types of passenger characteristics and their relation to behaviour are introduced in section 2.2. Information about these passenger characteristics can be collected using various methods, which are introduced in section 2.3. Finally, as the PDF encompasses a behavioural model, such models are introduced in section 2.4 in order to shed light on the usage of behavioural classes in the PDF. Finally, the chapter ends with section 2.5 , which summarizes the finding of the chapter and draws conclusions.

2.1 The passenger process

Passengers departing from, arriving or transferring at AAS have to go through a number of mandatory steps in the passenger process. Hence, the knowledge about this process can help predicting the next activity of a passenger. This is why the passenger process is a valuable source of information for a PDF. As such, the passenger process is discussed here.

Figure 2.2 shows the various steps that are involved in the passenger-centric operations at AAS, outlined by the white boxes. Due to European regulations, this passenger process is similar in all European airports. However, airports that are not subject to these regulations may have a different process than discussed here. Note that each step in the process can only be initiated once the previous process has been completed. The three grey boxes indicate the parts of the process that are specific to a certain type of passenger: departing, transferring, or arriving. In the remainder of this section, these three types of passengers and their process in the terminal will be discussed. Lastly, the difference between the process at the airport and similar locations such as train stations is discussed.

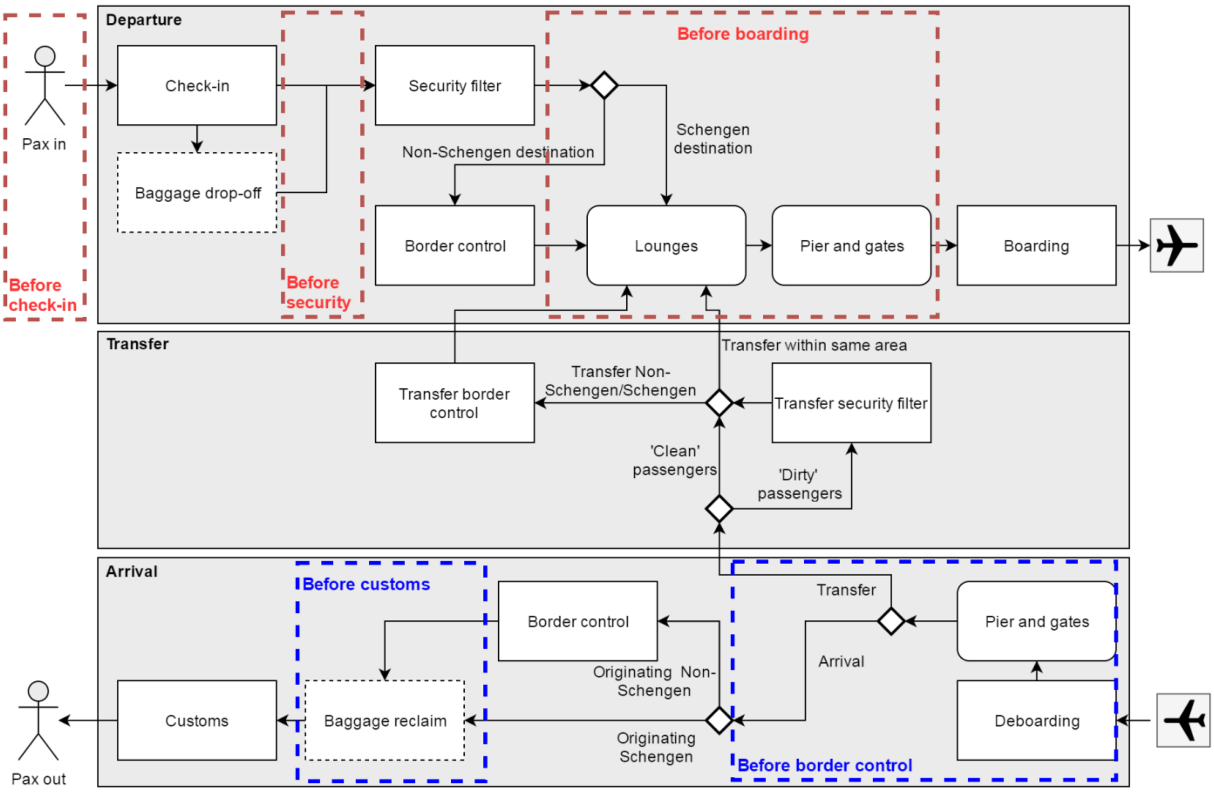


Figure 2.2: Passenger-centric operations at Schiphol, adapted from Grosmann (2016)

2.1.1 Departing passengers

Regarding activities and activity scheduling, departing passengers have the most complicated process out of the three types of passengers. This is because there are three main phases during their entire stay in the terminal (X. Liu et al., 2014). These phases are a result of the three main mandatory process steps during the passenger trip and the fact that lead times for these process steps are (to a certain extent) uncertain to the passenger. During the three phases, indicated by the red boxes in the figure, the passenger is free to perform discretionary activities not related to the passenger progress. The passenger decides when he will proceed to the next phase. Hence, the length of each phase can be designated as behaviour.

The first of the three phases is the *before check-in* phase, which is the time period that starts once the passenger arrives at the airport and ends once the passenger commences check-in (or baggage drop-off). Once the passenger has checked in, or in case the passenger has already checked in beforehand and has no baggage to drop off, he proceeds to the next main phase. This phase is the *before security* phase, which starts after check-in and ends once the passenger commences the security filter process (and passport control process, if applicable). The final phase is the *before boarding* phase, which starts after the security filter process and ends once the passenger starts the boarding phase.

As mentioned before, these three main phases are separated by mandatory steps in the passenger process. During the three phases, passengers can engage in discretionary activities as they please, although limited by the time remaining until boarding.

2.1.2 Arriving passengers

The process for arriving passengers is much shorter. All passengers arriving at AAS have to go through one mandatory passenger process, which is customs. In most cases, the lead time for this process is close to zero, given that a passenger has no goods to declare. For some high risk flights, elaborate baggage checks are in place, considerably increasing the time needed for customs. Passengers from a flight with a non-Schengen origin will also have to go through border control before they can reach the baggage reclaim halls. Due to the limited capacity of border control and the fact that often multiple non-Schengen flights arrive shortly after each other, the lead time for this process can be considerable.

Analogous to the three main phases for departing passengers introduced in section 2.1.1, the trip of arriving passengers in the terminal can be divided into two phases, indicated by the blue boxes in figure 2.2. The first phase can be defined as *before border control* which is the time between deplaning and starting the passport control process. The second phase is then the *before customs* phase. This phase starts after passport control and ends once the customs process is started. Note that passengers arriving from a flight with a Schengen origin also encounter the *before border control* phase. In their case, however, they do not have to show their passport, but walk through one-way doors once entering the baggage reclaim halls. Hence, the transition between these two phases marks a point of no return.

During the *before border control* phase, discretionary activities are possible, as passengers have access to the lounges at AAS. An exception here are so called ‘dirty’ passengers, who arrive from a flight that originated from a country with lower security standards compared to European standards and are hence regarded as unsafe. Dirty passengers from these flights arrive on a separate pier that does not grant access to the lounges. Consequently, these passengers can only directly proceed to border control.

During the *before customs* phase, passengers reside in the baggage reclaim hall in which only a very limited selection of discretionary activities can be undertaken as the amount of amenities is intentionally very limited in this part of AAS. Moreover, passengers in this part of their journey are mostly focussed on practical activities (such as informing themselves about their transportation away from the terminal) and an efficient execution of these activities (Schiphol Group, 2016).

2.1.3 Transferring passengers

The process steps that a transferring passenger goes through greatly depend on the properties of its arriving and departing flights. A dirty passenger will have to first proceed through a security filter before he can leave his arrival pier and proceed to the lounges. Clean passengers can go to the lounges immediately. A passenger transferring between Schengen and non-Schengen flights, or vice versa, has to go through border control.

2.1.4 Distinguishing aspects of the passenger process at airports

One might argue that the passenger process of passengers in an airport terminal is similar to other (passenger) processes and hence the behavioural classification developed in this thesis might just as well be applied to other situations, such as behaviour at a large event, or passengers in a train station. However, the main difference between pedestrian behaviour in airport terminals and other venues such as large events is the fact that the activity set of passengers in the terminal is largely mandatory. Whereas pedestrians at a big event can engage in any activity they want, passengers in an airport terminal have to go through a mandatory set of process steps. This notion is similar to passengers at a train station. Arriving at the station, passengers have to buy a ticket or check in with their public transportation card, they have to move towards the platform where their train will arrive and finally board the train once it arrives at the platform. However, due to being required to arrive at the airport well ahead of flight departure, passengers often have quite some time for other activities apart from the rigid mandatory passenger process.

It hence becomes clear from figure 2.2 that the passenger process is more involved compared to the process of train passengers and even more so compared to other venues. However, the rigidity and linearity of the passenger process does bring the benefit of making the activities that passengers will deploy in the terminal rather predictable. For example, it is certain that a passenger entering the queue for the security filter at a certain time will enter the lounge some moments later, depending on the current lead time for the security filter. On the other hand, passengers that have a lot of time to kill before their flight departs can engage in more discretionary activities, which are more difficult to predict. Nevertheless, such information based on the passenger process is very useful for creating forecasts and is hence included in PDFs, such as in the PASSME PDF as shown in figure 2.1.

2.2 Passenger characteristics and behaviour

The passenger process, discussed in the previous section, has shown that there are several mandatory steps for passengers going through an airport terminal, either departing, arriving, or transferring. Despite these mandatory steps in the process, the passenger has enough freedom to engage in discretionary activities and is free to decide when he proceeds to the next step of the process. Choices a passenger makes in this respect can be regarded as behaviour. Additionally, there are characteristics unique to the passenger that can be used to predict their behaviour within the context of behavioural classification.

The ensuing two subsections further elaborate on this. First, the definitions for the different types of characteristics that describe passengers and their behaviour are introduced. These definitions will be used throughout the rest of the report. Second, the relations between passenger characteristics and their behaviour are discussed, based on existing literature. From this, it becomes clear how passenger characteristics relate to passenger behaviour and, consequently, what this implies for behavioural classification.

2.2.1 Types of characteristics

In the previous parts of this report, various terms have been used to describe passenger behaviour and properties of passengers. When collecting data about passengers, this data can pertain to the passenger himself, such as his age or sex. Alternatively, it can pertain to his behaviour, such as the amount of time he chooses to spend in an area of the airport. For the sake of clarity, some definitions are introduced here. These definitions will be used throughout the rest of the report. We define two main types of characteristics: behavioural characteristics and passenger characteristics. Both of these characteristic types can be further specified into personal characteristics, process characteristics, and trip characteristics.

2.2.1.1 Behavioural characteristics and passenger characteristics

Behavioural characteristics pertain to the behaviour of a passenger. The definition is broad because the type of characteristics depends on the implementation in the PDF. Depending on the data used in the PDF, examples of such behavioural characteristics can be: walking speed, time spent in airport areas, and discretionary activities performed at the airport. In short, behavioural characteristics are those characteristics that are to be predicted by the behavioural classification. As such, the behavioural characteristics are typical for each behavioural class.

Passenger characteristics on the other hand describe the passenger himself. In the context of behavioural classification, these are the characteristics of a passenger based on which he can be placed in a behavioural class. Examples of passenger characteristics could be sociodemographic characteristics, flight number, and travel purpose.

Relating these two types of characteristics to the behavioural classification of this report, the following can be noted. When performing behavioural classification, the passenger characteristics of a passenger are known based on one or multiple data sources (i.e. sensors, app, airport database, or the airline database). Based on these passenger characteristics, a passenger is assigned to a behavioural class. This behavioural class represents the typical behaviour of that class, expressed in the values of the behavioural characteristics. These behavioural characteristics can then be used in the behavioural model of the PDF in order to model the passenger's behaviour.

2.2.1.2 Personal characteristics, process characteristics, and trip characteristics

Both of the aforementioned types of characteristics – passenger characteristics and behavioural characteristics – can pertain to various aspects of the passenger. For example: age is a passenger characteristic specific to the person. Another example is the flight destination. Although this is specific for a certain passenger, it does not say anything about the passenger itself, but rather about his trip. As such, the behavioural characteristics and passenger characteristics can be further categorised as personal-, process-, and trip characteristics.

Personal characteristics

Personal characteristics pertain specifically to one individual. This includes basic information about the person, such as age, gender and nationality. However, other information that is not inherently connected to the physical individual, such as group composition or carry-on baggage, is also regarded as personal characteristics.

Process characteristics

Process characteristics are related to the passenger process as described in section 2.1. More specifically, characteristics in this category describe the discretionary activities that a passenger has engaged in and the time he has spent on these activities.

Trip characteristics

Trip characteristics are related to the trip of the passenger from his origin airport to his destination airport and the locations at the airport associated with this trip. This includes, among others, gate number, airline, and flight number

2.2.2 Relation between passenger characteristics and behaviour

The premise of behavioural classification is that passenger characteristics are somehow related to the behavioural characteristics of a passenger. As such, the combination of passenger characteristics could explain the behavioural characteristics of a passenger. For example, one could intuitively expect that experienced travellers try to minimise their time at the airport and hence perform less discretionary activities. In order to verify the correctness of this premise, literature regarding passenger behaviour at airport has been reviewed. In addition, the information from literature allows setting expectations with respect to the results of behavioural classification.

In table 2.1, passenger characteristics are related to their behaviour at the airport. The relations described in this table are derived from the works of several authors. Their work is briefly described in the ensuing paragraphs.

Table 2.1: Passenger characteristics and their relation to behaviour

Passenger characteristic	Relation to behaviour	Reference
Age	Younger travellers are more likely to shop compared to middle aged travellers, who perform more facility use activities	X. Liu et al. (2014)
Airline type	Low-cost airline passengers are less likely to consume food or drinks at the airport	Castillo-Manzano and López-Valpuesta (2013)
Carry-on baggage	The higher the number of carry-on bags, the lower the chance of performing dining activities	X. Liu et al. (2014)
Check-in method	Passengers who have checked in online are less likely to perform discretionary activities before the security check	Kalakou and Moura (2015)
Education level	Travellers that have a higher education level are more like to perform inquiry activities	X. Liu et al. (2014)
Gender	Males are more likely to perform inquiry activities Females are more likely to perform shopping activities	X. Liu et al. (2014)
Group composition	Single travellers are less likely to shop, while travellers with children are less likely to engage in dining, shopping and facility use activities	X. Liu et al. (2014)
	Non-business travellers travelling in a group have a larger arrival time safety margin compared to individual non-business travellers	Tam, Lam, and Lo (2008)
	Passengers travelling in groups are more likely to engage in dining activities	Castillo-Manzano and López-Valpuesta (2013)
	Passengers travelling in groups that include children are less likely to engage in dining activities	Castillo-Manzano and López-Valpuesta (2013)
Income	High-income travellers are more likely to engage in shopping and dining activities	X. Liu et al. (2014)
Place of residence/ travel destination/ travelling company	Passengers who do not live in the city of the airport, are travelling to an international destination, and arrive at the airport accompanied by non-travellers, are more likely to perform discretionary activities before the security check	Kalakou and Moura (2015)
Total time at the airport	A higher total time at the airport increases the likelihood of food or drinks consumption	Castillo-Manzano and López-Valpuesta (2013)
	The time spent at the airport and consumption by passengers are positively correlated	Torres, Domínguez, Valdés, and Aza (2005)
Travel class	Economy class travellers have a longer waiting time in each of the three time phases than business class travellers	X. Liu et al. (2014)
Travel destination	Intercontinental passengers are more likely to consume food or drinks	Castillo-Manzano and López-Valpuesta (2013)
Travel experience	Frequent travellers are less likely to shop at the	X. Liu et al. (2014)

	airport	
	Frequent travellers and travellers who have planned their activities at the airport beforehand are less likely to perform discretionary activities before the security check	Kalakou and Moura (2015)
Travel purpose	Business travellers are less likely to perform discretionary activities before the security check	Kalakou and Moura (2015)
	Business passengers account for a larger safety margin when travelling to the airport by bus or taxi, compared to non-business travellers	Tam et al. (2008)

X. Liu et al. (2014) describe five types of discretionary activities: inquiry, dine, shop, wait, and facility use. These types of activity were used in a nested logit choice model that predicts passenger activity scheduling behaviour based on passenger characteristics. The authors looked only at departing passengers and distinguished three time periods, as presented in section 2.1.1. The authors drew several interesting conclusions with respect to the relation between passenger characteristics and behaviour, as shown in table 2.1.

Kalakou and Moura (2015) constructed a discrete choice model for behaviour of departing passengers before the security filters, based on survey data collected at Lisbon Portela airport. Tam et al. (2008) modelled passenger behaviour based on stated and revealed preference for the situation of Hong Kong International Airport. They modelled access mode choice and arrival safety margin choice, which is the extra time passengers allow between their preferred arrival time at the terminal and the expected arrival time for their chosen access mode.

Also referenced in table 2.1 is the work of Castillo-Manzano and López-Valpuesta (2013), who researched the behaviour of passengers with respect to catering facilities at airports, based on surveys at eight different airports, yielding a sample of as many as 37,000 passengers. They concluded that eating and drinking are the most-performed discretionary activities at airports. The primary motivator for these activities is waiting time. Furthermore, as many as 22 variables with respect to passengers and their trip were found to be significantly related to the likelihood of consuming food or drinks. Seventeen of these variables were even significant at the 1% level.

Apart from the effect of personal characteristics, the airport itself also has an effect on passenger behaviour. In large airports, travellers are less likely to wait before checking in and perform discretionary activities instead. Additionally, at small airports, travellers are less likely to perform shopping activities, although this may be because there are few shopping facilities present in any case (X. Liu et al., 2014). Additionally, numerous studies have found moderating effects of time-pressure on shopping behaviour (Lin & Chen, 2013). Conversely, the time spent in an airport and the consumption by passengers are positively correlated (Torres et al., 2005).

As shown in the previous paragraphs, there are several established relations between passenger characteristics and passenger behaviour. However, it should be noted that all referenced works used surveys. Consequently, it should be taken into account that not all passenger characteristics mentioned in table 2.1 can be acquired within the context of an operational PDF, i.e. using the information sources as shown at the start of this chapter in figure 2.1. For example, it is not possible to acquire a person's education level or income without specifically asking this person.

Although the several references that have been used mostly researched different passenger characteristics, there is some overlap between them. For these characteristics, it can be noted that the different sources mostly agree:

- Group travellers are more likely to engage in dining activities, except for groups with children.
- Experienced travellers are less likely to perform discretionary activities.

- Business travellers perform fewer activities before the security check and have a larger arrival safety margin; for these passengers it is valuable to not miss their flight.
- Passengers spending more time at the airport are more likely to consume food or drinks.

Overall, it can be noted that, out of all characteristics, group composition influences the most aspects of behaviour. This characteristic is related to behaviour regarding shopping, dining, performing non-discretionary activities, and the time spent in various areas of the airport. As such, this behavioural characteristic could be important to have available for behavioural classification. However, also the other passenger characteristics shown in table 2.1 should ideally be available for behavioural classification as they have a relation to one or more aspects of passenger behaviour.

2.3 Collecting passenger data

The first sections of this chapter have discussed the passenger process and the relations between passenger characteristics and passenger behaviour. The terms 'behavioural characteristic' and 'passenger characteristic' have been introduced, indicating the different aspects of a passenger and his behaviour in relation to behavioural classification. Based on literature, the relations between several of these two types of characteristics have been shown. Because of their shown relation, such characteristics should be available for behavioural classification. However, some of the characteristics mentioned in section 2.2 can only be obtained by means of surveys, which would not be practical in an actual implementation of behavioural classification in combination with a PDF. Moreover, the set of characteristics of the previous section is non-exhaustive; several other characteristics could be useful for behavioural classification.

As such, this section introduces several other characteristics, and the data sources that can be used to acquire these characteristics. The remainder of this section is divided into two subsections. The first subsection gives a general overview of passenger data collection methods and other data sources, ranging from tracking of mobile devices to acquiring data from the airline database. This gives an idea about how various characteristics can be collected. The second subsection introduces a number of characteristics and relates them to a specific data source, focussed on the case of AAS.

2.3.1 Passenger data collection methods and other data sources

In order to be able to segment passengers into different behavioural classes, data about their behaviour and characteristics have to be available. This subsection introduces various methods that can be used to collect data about such characteristics. This includes currently existing and proven methods, as well as methods that are still being developed.

Pedestrian data collection can be divided into four main data collection methods (Millonig, Brändle, Ray, Bauer, & van der Spek, 2009): Position Determination Technologies (PDT), video-based data collection, observations, and surveying. PDT can be used to determine the position of individuals, which can provide valuable information about their location and activities in the terminal and the amount of time they spent at these locations. Video-based data collection uses video footage from cameras to determine characteristics about the motion of people and their individual attributes. Observations and surveying are offline and manual methods that are most suitable to provide information about the decision process and factors influencing these decisions. In addition to these four main data collection methods, other sources of data collection specific to the airport are also discussed. Fitting with the concept of the PASSME PDF as presented in figure 2.1, these are the airline database, airport database, and mobile app.

2.3.1.1 Position Determination Technologies (PDT)

Position determination technologies (PDT) is a general term for technologies that can provide the location of an object or person. There are several traditional PDTs, such as GNSS or cell-based positioning. Additionally, there are newer techniques such as position determination based on Bluetooth or Wi-Fi signals. These are especially interesting due to the high increase in the percentage of smartphone owners worldwide in the last couple of years (Poushter, 2016). In the ensuing, the aforementioned PDTs are further elaborated on with respect to their usability in an airport environment.

Satellite and cell-based positioning

One of the best known PDTs is probably Global Navigation Satellite Systems (GNSS). GNSS is based on a system of non-geostationary satellites that broadcast a signal. There are a few of these satellite systems in place, of which the American GPS is the best known. Additionally, there is the European GALILEO and the Russian GLONASS system. GNSS receivers compute their location based on the signals they receive from the satellites. This signal is one-way; hence there is no communication between the receiver and the satellite. As a result, only the receiver is aware of its location. Major benefits of GNSS are the fact that they are free to use and that GPS-devices are relatively low-cost**.** Position determination is high frequency and quite accurate in good conditions. However, signal obstructions are detrimental to the performance. This means that GNSS is not suitable for indoor use (Harle, 2013; H. Liu, Darabi, Banerjee, & Liu, 2007; Millonig et al., 2009).

Another PDT is cell-based positioning, which is based on mobile telecommunication networks. This technique determines location based on the antennas that the mobile device is connected to. The technique is suitable for indoor use such as in an airport terminal. However, in this case the resolution is quite low. On top of that, data collection can be done client-side or server-based. In the first case, the client-device determines its location and is thus not directly available to the researcher. In the case of server-based collection, the cell network provider determines the location of the connected devices. In this case, however, privacy is a concern that should be addressed (Millonig et al., 2009). Generally, the accuracy of cell-based positioning is low (H. Liu et al., 2007), although there are examples of increasing accuracy by deploying base stations inside buildings (Otsason, Varshavsky, LaMarca, & De Lara, 2005).

RF-positioning

Bluetooth, a short distance radio communication technology, can also be used as a means for position determination. Bluetooth tracking is based on the tracking of the unique MAC address of the device. This tracking can be done in two ways; actively or passively. With passive tracking, a fixed network of Bluetooth devices scans for other Bluetooth devices in the vicinity. Found Bluetooth devices are registered, yielding the approximate location. Active tracking works the other way around; the client device scans for a network of Bluetooth beacons. Based on the scanned devices and the information about the network infrastructure, the client device determines its location. Similar to the case of client-side cell-based positioning, the position data is not directly available to the researcher, unless explicitly shared. Position determination using Bluetooth is relatively accurate, with an accuracy of five to ten metres (Millonig et al., 2009). Specific implementations of Bluetooth tracking even mention accuracies better than five metres (H. Liu et al., 2007). Due to its dependence on a network of devices, this position determination technology is suitable to use indoors. A downside of this method is the fact passengers need to have Bluetooth enabled. It turns out that the percentage of people that have this is as small as 5% (Millonig et al., 2009). More recent studies, however, mention much higher numbers, such as 34% in the case of supermarket shoppers (Phua, Page, & Bogomolova, 2015).

Wi-Fi tracking is a technique similar to Bluetooth tracking. This technique also relies on a network of sensors and the registration of MAC addresses and determines the position of the Wi-Fi device based on the trilateration of the time of arrival of the signal (Vorst et al., 2008). The accuracy of this system again depends on the network, but an accuracy of up to three metres can be expected (H. Liu et al., 2007; Vorst et al., 2008).

Another possibility for RF-positioning is radio frequency identification (RFID). This technology relies on RFID tags that have to be carried by the subject. These tags can be either passive or active (Millonig et al., 2009)**.** Active RFID tags are powered by a power source and emit a signal. Similar to Wi-Fi and Bluetooth tracking, this signal can be picked up by a network of RFID readers. Based on the strength of the received signal, the location of the tag can be determined via trilateration. Passive RFID tags are powered by the signal from the RFID reader, which means that the range between the tag and the reader is very limited. Another approach presented by Vorst et al. (2008) equips the subject with an RFID reader and places stationary RFID transponders with a known location in the environment. The location is then determined using a particle filter. The authors report an accuracy of about 0.4 metres (Vorst et al., 2008), while the aforementioned passive and active RFID tracking have an accuracy of respectively a few centimetres and approximately 10-100 metres (H. Liu et al., 2007; Millonig et al., 2009). However, in any case a major limitation of this method remains the fact that subjects have to be equipped with a specific device to be traceable.

2.3.1.2 Video-based data collection

Imagery from video cameras can provide a lot of information about crowds. Computer vision techniques are employed to extract information from these images, given that the camera setup is appropriate. There are three main types of information that are extracted from video (Millonig et al., 2009; Zhan, Monekosso, Remagnino, Velastin, & Xu, 2008): crowd density, recognition and tracking.

Crowd density measurement can count the number of people in a specific scene, and consequently estimate the density in that scene. Density measurement techniques are divided in three groups (Silveira Jacques Junior, Musse, & Jung, 2010): pixel-based analysis, texture-based analysis and object level analysis.

Recognition techniques recognise individuals in scene. Current commercial implementations of facial recognition are even able to perform advanced facial analysis, yielding detailed individual information, such as mood, age, sex, ethnicity and clothing style (Sightcorp, n.d.).

Tracking of individuals relies on the recognition of individuals within the image across multiple time instances. Often, computer vision techniques are enhanced with a model to help predict the movement of individuals (Zhan et al., 2008). Other techniques in literature are able to detect groups and human interaction (Tran, Gala, Kakadiaris, & Shah, 2014), or detect specific objects in a scene (Hu, Tan, Wang, & Maybank, 2004). Zaki and Sayed (2014) present a method for automated walking gait analysis that is able to estimate individual pedestrian's age category and gender.

3D vision techniques such as LIDAR can provide more detailed information about the environment and objects with respect to distances, increasing tracking accuracies. Techniques like these scan the environment and determine the distance to objects and the environment, resulting in a point cloud.

2.3.1.3 Manual observations and survey techniques

Apart from sensors and other devices to collect pedestrian data, data can also be collected by hand. Methods like this include observations and survey techniques. Obviously, these methods are not suitable to use directly as an input in a passenger demand forecast system as this data is not real-time. However, the available methods are mentioned here as they can be useful to acquire an initial data set based on which a preliminary classification can be made.

Observations

The first manual method is observation. With observation, a subject is followed and his behaviour is observed. Millonig et al. (2009) distinguish three main types of interest for pedestrian data collection. First, direct (reactive) observation in which the observer identifies himself and the goal of his observations. The second type is unobtrusive observation in which the observer does not introduce himself to the studied subject. The subject is unaware that it is being observed, hence reducing the risk of the subject altering his behaviour. This type of observation can be done on-site, but can also be based on video recordings. The third type is participatory observation. With this type of observation, the observer participates in the phenomenon that is being observed, giving him a closer look. Observations are generally regarded as most suitable for explorative and descriptive research due to its flexible nature.

Survey techniques

Survey techniques are the second manual method. According to Millonig et al. (2009), survey studies are one of the most important data sources for research with respect to influence factors on human route decision making. Using survey techniques, subjects can be asked about their behaviour. However, respondents tend to idealize themselves when asked to participate in a survey, hence always leading to a certain degree of inaccuracy. Millonig et al. (2009) differentiate three main survey techniques: questionnaires, interviews and trip diaries. Questionnaires can easily be distributed amongst a large number and hence lead to a relatively large standardised data set. However, the quality of the data can fluctuate as participants can interpret questions differently, or if the given answering options do not properly reflect the participants' opinions. Lastly, the answering options can also influence the participants' response (Millonig et al., 2009).

The second technique, interviews, can be conducted in either a structured or non-structured fashion. In the first case, the interview questions are standardised for each interview. In the latter case, the questions are not standardised. Interviews yield highly detailed responses, but the sample size is generally rather low due to the method being quite time intensive. Additionally, processing interviews is quite involved as it requires categorisation of the responses.

Trip diaries contain the activities, location and duration of the behaviour of the participant. Participants fill in these diaries themselves, either as a recall diary, or as a self-administered diary. Recall diaries are made ex post and consequently rely on the memory of the subject. Self-administered diaries, on the other hand, are filled in real-time. Both methods require a lot of effort of the participant and can give results of varying quality (Millonig et al., 2009).

2.3.1.4 Airline database

Airlines store their booking information in their own data systems. Hence, this database contains information about the name of a passenger, and their complete travel itinerary. In addition, a lot of airlines offer loyalty programs that could provide additional personal characteristics of the passenger, such as age and gender. As such, the airline database could theoretically provide a lot of personal and trip passenger characteristics.

2.3.1.5 Airport database

The airport database contains operational information with respect to flights. This includes departure time, departure gate, check-in desks et cetera. Hence, the airport database mostly contains operational information.

2.3.1.6 Mobile app

An airport app, such as the one for AAS, provides passengers with information about their flight and the airport amenities. This way, travellers can easily gather information on their mobile device. The Schiphol app also offers position determination. An airport app like this could be enhanced to let the passenger opt-in to share their location and personal characteristics.

2.3.2 Data sources at AAS and data fusion

In subsection 2.2.2, various passenger characteristics have been introduced that have a proven relation with passenger behaviour. Such characteristics should hence be available. However, there are many more characteristics that are not discussed in literature, but which could be used for behavioural classification. Therefore, an extensive overview of 37 passenger and behavioural characteristics has been compiled. Due to its large size, spanning multiple pages, it has been included in Appendix A. However, hereafter the contents of the overview are described on a global level.

For each characteristic in the overview, the following information is given:

- **Type of characteristic** indicates whether the characteristic is a passenger characteristics or a behavioural characteristic.

- **Category of characteristic** gives the category of the characteristic (according to the definition of section 2.2.1.2), which can be personal, process, or trip.
- **Value** indicates the type of value of the characteristic; i.e. categorical or numerical.
- **Data source** indicates what the source for the characteristic could be, based on the technologies mentioned in the previous subsection. It is also indicated whether the source is presently available, could possibly be implemented at AAS, or is still in development as a technology. In addition, it is indicated if the information from the data source is linked to a specific passenger, or if it is not linked to a specific passenger and hence some sort of data fusion is required to be able to do so.
- **Comments** further clarify what has been indicated in the data source columns.

Out of the total of 37 characteristics, eight are behavioural characteristics. Some examples of these are walking speed, activities engaged in the terminal, and the time spent in each step of the passenger process. The remaining 29 characteristics are passenger characteristics. Examples of these are the age, sex, nationality group composition, and the type of carry-on baggage (if any).

For some characteristics, there are already systems in place at AAS that could provide such information. One of such systems is Bliptrack, which is an RF-positioning system that uses the Wi-Fi and Bluetooth signals of passengers' devices to track them through the terminal. Although this could theoretically supply quite detailed location information, the current implementation of the system is focussed on lead time prediction and the resolution is hence quite low. Another example of such as system is the Self-Service Boarding Pass Check (SSBPC). At the SSBPC, passengers scan their boarding pass in order to get access to the security filters. The SSBPC can provide information about the time a specific passenger has entered the security process and information related to the boarding pass such as the name of the passenger, his flight, gate number, et cetera.

For some of the passenger characteristics mentioned in the appendix, subsection 2.2.2 has already shown that these are related to passenger behaviour and should therefore be available for behavioural classification. However, although for the remaining characteristics there is as of yet no proven relation between the characteristic and passenger behaviour, these should ideally be available for behavioural classification as this maximizes the amount of information used in the behavioural classification. Based on the importance of the characteristics in the classification results, it can be decided if these characteristics are indeed necessary for classification.

Out of the 37 characteristics, 27 could be acquired using a current source of data. In spite of this, there are two main challenges to be overcome. First, the mentioned sources of data are not coupled. Let us illustrate this with an example. Passenger X checks in at a check-in desk. Now, the airline database 'knows' when X has checked in, his amount of baggage, flight number et cetera. Before X checked in, his phone has already been registered at multiple Bliptrack beacons. However, while the airline database has X's name and certain characteristics, the Bliptrack system registers X's device identification and certain characteristics. The system does not register that X's device actually belongs to passenger X. Hence, both systems provide characteristics about the passenger, but one is related to the passenger's name, and the other to the passenger's device. Consequently, the challenge is to combine these two. This example can also be related to the other data sources that have been mentioned. As such, it can be concluded that data fusion of such sources presents a challenge. A data set following from such data fusion will lead to a very extensive overview of passenger and behavioural characteristics on an individual level. The privacy of the passenger should thus be insured, for example by anonymizing the data. Privacy is hence the second challenge.

2.4 Passenger behavioural models

The result of behavioural classification of this report is to be used in the behavioural model of the PASSME passenger demand forecasting system. Though the specifics of this model are not clear yet, it is to be based on pedestrian behavioural models. Therefore, a brief overview of such models is given in this section.

2.4.1 Terminology

From literature about pedestrian behavioural models, it becomes apparent that there are multiple terms used to refer to models that attempt to recreate pedestrian behaviour. Terms such as 'crowd simulation', 'pedestrian simulation', 'pedestrian behaviour model', 'behaviour modelling', 'crowd behaviour model' and 'crowd dynamics model' are commonly used in literature. Furthermore, the use of the word 'crowd' is also subject to different interpretations (Duives, Daamen, & Hoogendoorn, 2013), some referring to at least two individuals, others referring to thousands of individuals. Ali, Nishino, Manocha, and Shah (2013) define it as "any collection of individuals or pedestrians where behaviour of one individual is influenced by the other". The chosen term is usually not related to the type of modelling that has been chosen, although authors using 'pedestrian behaviour' or a similar term generally describe agent-based models that are focussed on recreating pedestrian behaviour rather than crowd phenomena.

In this report, the term 'behavioural model' is used to indicate models that model any type of pedestrian-, or, in this case, passenger behaviour. This behaviour can be on any level; strategic, tactical or operational.

2.4.2 Levels of pedestrian behaviour

Literature on behavioural modelling generally defines three levels of pedestrian behaviour (Daamen, 2004; Hoogendoorn, 2001; Hoogendoorn & Bovy, 2004). An overview of these levels is given in figure 2.3. A combination of the three interacting levels of behaviour in a single model would provide a model that covers the entire scope of pedestrian behaviour. Nevertheless, the scope of many models in literature is on specific cases such as evacuation scenarios or the recreation of crowd phenomena, which both often focus only on the operational aspect of behaviour. Consequently, the focus of these models is on the tactical and operational level. In these cases, the strategic level is regarded as an input to the model, rather than as a part of the model itself (Abdelghany, Abdelghany, & Mahmassani, 2016; Daamen, 2004; Hoogendoorn & Bovy, 2004).

2.4.2.1 Strategic level

The first and highest level is the strategic level. On this level, a set of activities is chosen and ordered. This set of activities contains discretionary and mandatory events. In the case of Schiphol, a discretionary event could be buying a coffee, while an example of a mandatory event is passing the security filter. A part of the activity set has been formed a priori, i.e. before the pedestrian has entered the terminal. However, certain activities can also be added to the activity set based on the strategic level, e.g. deciding to buy a coffee when there is time left to do so. Though there are models available that include the strategic level, not much work is directly related to pedestrians. Additionally, in behavioural models, the strategic level is often regarded as exogenous to the model.

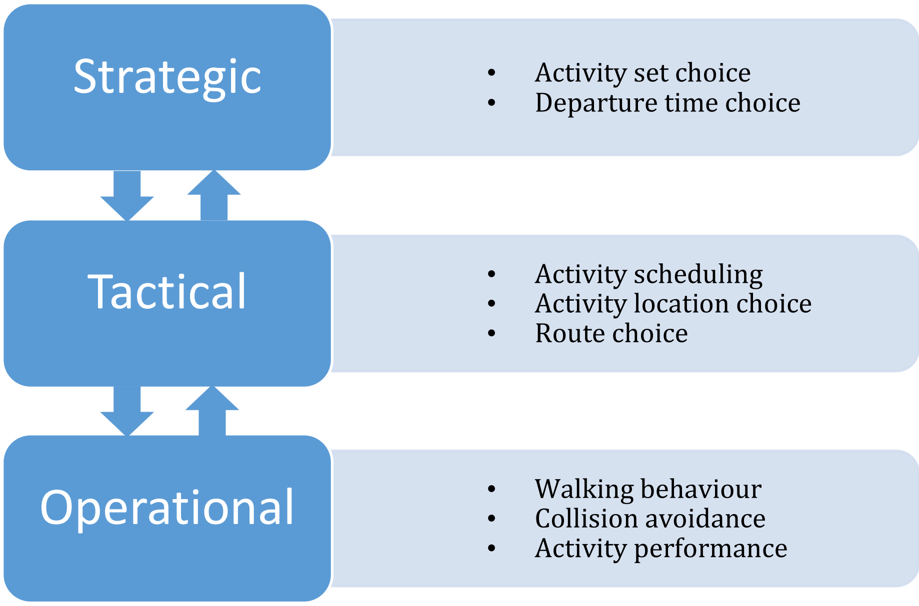


Figure 2.3: Levels of pedestrian behaviour

2.4.2.2 *Tactical level*

The second level is the tactical level. This level pertains to the short-term decision of passengers. Based on the activity set from the strategic level, activities are scheduled and the activity location is chosen (Hoogendoorn, 2001). The route to these activities is also chosen at this level. Activities in the activity set may be reordered or removed

The input from the strategic level, i.e. the set of chosen activities, has to be scheduled on the tactical level. This (ordered) list of activities, containing discretionary and mandatory events, will be ordered based on the amount of time available. Discretionary events may be removed from the set if there is not enough time available to perform this activity (Daamen, 2004).Furthermore, activities may be added, should the activities in the activity list not fulfil the time available to the passenger. In addition to the activity, the location of the activity is also chosen.

During route choice, a path towards the chosen goal has to be chosen. Various methods can be used to generate route choice sets, see for example (Ali et al., 2013). Based on the generated choice set, a route is chosen, often using either a probit or logit model (Daamen, 2004).

2.4.2.3 *Operational level*

The third level is the operational level, which contains walking behaviour. Behaviours involved in this level include movements towards the goal defined on the tactical level, collision avoidance with other pedestrians or static objects such as walls.

2.4.3 Types of behavioural models

Behavioural models consist of three main categories: macroscopic models, mesoscopic models and microscopic models (Daamen, 2004), though these categories mainly apply to the models on the operational and tactical level. Microscopic models model crowds as discrete individuals that behave and interact according to their own behavioural rules (Zhan et al., 2008). The behaviour of the crowd as a whole is hence the result of all individuals in the model. Conversely, the crowd is modelled as a whole and represented as a density in macroscopic models (Schadschneider, Klüpfel, Kretz, Rogsch, & Seyfried, 2009). Crowds are often modelled flow-based or particle-based, respectively adhering to fluid dynamics laws or physical laws (Kountouriotis, Thomopoulos, & Papelis, 2014; Schadschneider et al., 2009). Mesoscopic models combine microscopic and macroscopic properties and are based on a statistical distribution of the states of pedestrians in the model (Cristiani, Piccoli, & Tosin, 2014).

A special category of microscopic models are agent-based models. The envisioned PASSME PDF will use an agent-based model. In such models, each pedestrian is modelled as an individual agent. These agents each possess certain characteristics, such as top speed, acceleration, and how inclined they are to follow others. Agents behave and interact according to their characteristics and the governing rules of the model. An agent-based model can either have a homogeneous or heterogeneous population. In case of a homogeneous population, each agent in the model has the same properties. In a heterogeneous population, agents have different properties. This means that agents react differently, for instance because they have a different desired following distance, or because an agent is a follower rather than a leader. Hence, a behavioural classification methodology such as developed in this thesis could be used to populate an agent-based model with a heterogeneous population.

2.4.4 Implications for behavioural classification

As mentioned before, the behavioural classes resulting from behavioural classification form the input for a behavioural model. Consequently, the type of behavioural model imposes requirements on the output. This means that the input parameters for the model that are desired should be available for classification. Following the definition of characteristics from 2.2.1, these input parameters are hence the behavioural characteristics. Classes resulting from behavioural classification each have specific values for the behavioural characteristics. A passenger modelled in the behavioural model is then given the model input parameters corresponding to the behavioural class he was assigned to.

2.5 Chapter conclusion

In this chapter, the context of behavioural classification has been further explained. It has been shown that, as part of the PASSME project, a new PDF-system is being developed that combines data from multiple sources of (sensor) data. Based on this data, and the passenger process, a demand forecast is made. The behavioural classification of this thesis forms the input for the behavioural model of the PDF. The different aspects of the PDF have been discussed in this chapter: the passenger process, characteristics and their relation to passenger behaviour, sources of passenger data, and behavioural models.

With respect to the passenger process, the different processes for departing, transferring and arriving passengers have been shown. Although the passenger process is quite rigid, including several points of no return, passengers often have enough time to perform other optional activities. Based on the differences between the processes of the three types of passengers, it can be concluded that these three types should be treated separately.

Two important definitions regarding the characteristics of passengers have been presented. The first is behavioural characteristics; these describe the behaviour of the passenger and form the input of a behavioural model in the PDF. Behavioural characteristics are those characteristics that follow from the behavioural classification. Each behavioural class has its own values for the behavioural characteristics. The second definition is passenger characteristics; these are the characteristics that describe the passenger. These are collected using various (sensor) data sources. Based on their passenger characteristics, a passenger is assigned to a behavioural class.

Literature has shown that there are several established relations between passenger characteristics and behaviour. For fourteen passenger characteristics, it has been shown that these are related to one or more aspects of behaviour: activities performed at the airport, likelihood of consumption of food or drinks, time spent in the different phases of the passenger process, and time margin with respect to the arrival time at the airport. Because these characteristics are related to behaviour these should preferably be available for behavioural classification.

In addition to the passenger characteristics for which a relation with behavioural characteristics has been established, there are many more possible characteristics that could be used. An overview of 37 possible characteristics with their associated possible data source, based on the case of AAS, has been presented. This has introduced two challenges with respect to the collection of passenger and behavioural characteristics in an airport environment. First, while there are many possible sources of data in an airport environment, these sources are mostly unrelated. This means that data fusion is necessary in order to combine the data sources so that data is available on an individual level. The second challenge is assuring privacy, as collecting such a detailed set of data on an individual level can compromise passenger privacy.

The PASSME PDF encompasses a behavioural model, for which the behavioural classes should form the input. As the specifics of this behavioural model are not yet determined, some main principles of behavioural models have been presented. Roughly, passenger behaviour can be segmented into three levels: strategic, tactical, and operational. On the strategic level activities are chosen. On the tactical levels, activities are (re-)ordered and the route to these activities is chosen. The operational level mainly pertains to walking behaviour, such as collision avoidance. Addressing the sub-question for the modelling block of the thesis, it can be concluded that a behavioural classification could interface with any of the three levels of behaviour, as long as data with behavioural characteristics for these levels are available. Based on these behavioural characteristics, behavioural classes can be made, which may be used in a behavioural model. Then, 'new' passengers may be classified to one of the behavioural classes based on their passenger characteristics.

In the next chapter, a framework for behavioural classification will be introduced. This framework is based on the information as set out in this chapter. It classifies passengers into behavioural classes based on their passenger characteristics. The classes that follow from the classification carry behavioural characteristics specific to that class. For the framework, it is assumed that a data set containing passenger and behavioural characteristics based on the data sources as mentioned in this chapter is available.

3

Clustering and Classification Framework

In the previous chapter, various aspects of passengers, passenger behaviour, and the passenger process have been discussed. Additionally, methods for data collection with respect to the behavioural and personal characteristics of passengers have been discussed. Furthermore, an overview of various pedestrian behaviour models has been given.

While the previous chapter focussed on the sensing and modelling blocks, this chapter shifts the focus to the processing block. It introduces and tests the framework that will segment passengers into behavioural classes and establish rules to classify new observations to a specific behavioural class. This framework consists of two parts. The first part creates the behavioural classes based on behaviour. The second part creates the classifier that can classify passenger into one of the behavioural classes. For both of these parts, several possible techniques that could be used will be discussed, and one of these techniques is chosen and implemented. The techniques chosen for both parts are implemented. This integrated framework will then be used on a real data set in chapter 4, which will provide some insight into the possible performance of the framework.

This chapter is structured as follows. The chapter starts with section 3.1, wherein the requirements for the framework are set out. Section 3.2 then presents the overall framework. The subsequent sections 3.3 and 3.4 further elaborate on the two parts of the framework: clustering and classification. Section 3.5 discusses the integration of these two main parts and presents the implemented process as a whole. The chapter ends with a conclusion in section 3.6.

3.1 Framework and data requirements

The previous chapter has introduced various aspects related to the different parts of the PASSME PDF-system. The passenger process, various sources of data, behavioural models, and the difference between passenger attributes and behavioural attributes were discussed. Based on this information, a framework for behavioural classification will be introduced later in this chapter. This framework performs two tasks. First, it segments passenger behaviour into, and thereby defines, behavioural classes. Second, it establishes classification rules to assign passengers to such a behavioural class.

The framework is aimed at using (sensor) data as input. The output of the framework is aimed at usage in a behavioural model (of the PASSME PDF). Because of this, there are certain requirements to be set with respect to the inputs and outputs of the model. Additionally, a certain performance for behavioural classification is required. All of these requirements will be discussed in the ensuing subsections.

3.1.1 Framework input

The input for the framework is data about passengers. To be able to create a classification, both of the two types of characteristics introduced in section 2.2.1.1 are required in the data: passenger characteristics and behavioural characteristics. The framework should use the behavioural characteristics to form behavioural classes. The passenger characteristics should be used to assign passengers to a behavioural class. Based on the various types of possible input data, the framework should be able to handle both categorical and numerical variable types.

Because the behavioural classification is on an individual basis, the data provided to the framework should be formatted on an individual basis. This means that every object in the data represents an individual passenger. To train the framework, both passenger characteristics and behavioural characteristics have to be available. From an operational viewpoint, this means that both types of data should be collected using sensors or other data sources, such as mentioned in section 2.3. This data should then be aggregated into a data set that contains many observed passengers. Once the framework has been trained using the aggregated training data, newly observed passengers should be classified solely based on their passenger attributes.

3.1.2 Framework output

The framework should have two main outputs. First, it should output a classifier that can assign passengers to a behavioural class. Second, it should output the characteristics of each behavioural class. These class characteristics are the typical values of the behavioural attributes of each behavioural class. The attributes in the behavioural classes depend on the types of attributes that have been used as input to the framework. As such, the behavioural parameters that are desired in the outputted behavioural classes should also be available as input to the model. Furthermore, it is important to be able to assess the performance of both the behavioural classes that are formed and the subsequent classification. Performance metrics that reflect this performance are hence needed.

3.1.3 Framework performance

When used with actual data, the performance of the framework should be assessed in order to determine whether the results of behavioural classification are satisfactory. With respect to the behavioural classes, it should be tested whether there are indeed classes present in the data. As such the performance in case of more than one class should be better than the performance of a case with one class, which indicates that there are indeed classes in the data.

Furthermore, classification also requires a certain performance threshold. However, such a threshold is hard to define as the performance that can be acquired greatly depends on the type of data used. There are various performance measures available for classification, each representing different aspects of the performance. Even though this limits the ability to set a specific performance requirement, it can be said that the performance of the framework should at least be higher than the performance that would have been achieved by randomly assigning passengers to a behavioural class.

3.1.4 Requirements summary

Summarizing the requirements of the framework with respect to data input, output, and framework performance, yields the following list of requirements:

Framework input:

- The framework requires and uses behavioural attributes and passenger attributes.
- Data used in the framework contains information per individual passenger.
- The framework should be suitable for categorical and numerical attributes.
- Behavioural classes are formed based on the behavioural attributes in the data.
- Classification rules are formed based on the passenger attributes.

- In order to train the framework (i.e. creating the behavioural classes), passenger attributes as well as behavioural attributes are to be provided. Once trained, the framework should classify passengers solely based on their passenger characteristics.

Framework output:

- The framework should have two main outputs:
  - The characteristics of each behavioural class.
  - A classifier to assign passengers to a behavioural class.
- Attributes present in the behavioural classes also have to be available as input to the framework.
- Performance metrics that allow validation of the behavioural classes and the classification are required.

Framework performance:

- There should indeed be classes present in the data, i.e., the performance of using more than one class should be higher than using just one class.
- The classification performance should at least be better than the performance that would be achieved by randomly assigning passengers to a behavioural class.

3.2 The passenger clustering and classification framework

Based on the requirements laid out in section 3.1, a framework for behavioural classification has been designed. The framework receives passenger characteristics and behavioural characteristics and performs two main tasks:

1. **Clustering:** Based on the behavioural characteristics of passengers, behavioural classes are formed.
2. **Classification:** Based on the passenger characteristics of passengers, passengers are assigned to a behavioural class.

The framework is further elaborated on in the ensuing subsections.

3.2.1 The framework

Figure 3.1 presents the framework that segments input data into various behavioural classes and then creates classification rules. In the ensuing, this is referred to as the clustering and classification (CC) framework. The CC-framework forms the basis of all further work in this thesis report.

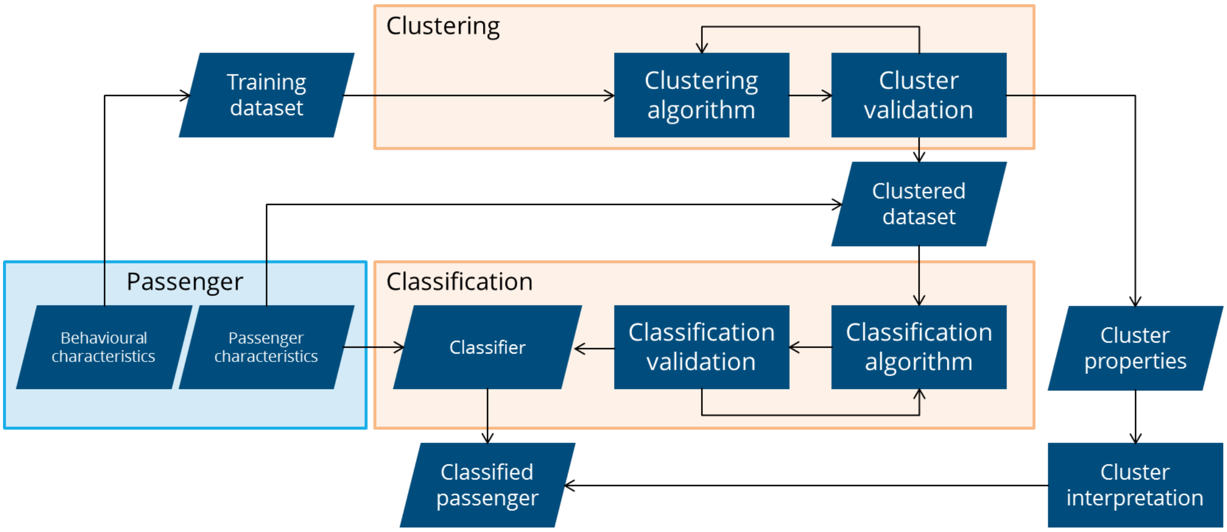


Figure 3.1: Passenger Clustering and Classification (CC) framework

The framework consists of two main parts: clustering and classification. The input for the framework is passenger data, i.e.: the behavioural characteristics and passenger characteristics as they have been previously defined. In the remainder of this subsection, the CC-framework will be explained based on the input, clustering, and classification.

3.2.1.1 Input: passenger characteristics and behavioural characteristics

The input for the framework is shown in the blue block on the left of the figure. This input consists of data about the passenger, which are the behavioural characteristics and passenger characteristics, as introduced in section 2.2.1. Further on in this thesis, survey data will be used as the passenger data. However, in an operational scenario, this data consists of the information collected through sensors, the airport database, and airline database.

The behavioural characteristics in the input data are used to form behavioural groups using clustering. Note that behavioural characteristics are only required when (new) behavioural groups are to be formed; once the classifier has been constructed, only the passenger characteristics of a passenger are needed to perform behavioural classification. Thus, the behavioural characteristics are predicted by the behavioural class to which the passenger is classified based on its passenger characteristics.

3.2.1.2 Clustering

Behavioural classes are formed in the clustering part at the top of the framework. Because the goal is to find groups of passengers with respect to behaviour, clustering techniques are used. These techniques can group a set of data into homogeneous subsets. Because clustering is unsupervised, no pre-classified training data have to be provided to the algorithm. Consequently, the behavioural classes that are formed are solely based on the behavioural attributes that are provided. Contrary to, for example, a discrete choice model with multiple classes, cluster analysis does not require estimating a model. This takes out the risk of trying to fit a model that cannot properly describe the data. Nevertheless, cluster analysis could also erroneously imply the presence of groups in data. It is therefore important to assess the validity of the clusters that were found, as shown in the framework.

The clustering part of the framework takes the training data set as input. Based on the behavioural characteristics, behavioural classes are formed by the clustering algorithm. There are two main outputs from clustering. The first output is the input data set augmented with information about which cluster each object in the data has been assigned to. The second output contains the characteristics of the clusters that have been formed, i.e., the passenger characteristics and behaviours that are specific to that cluster. These clusters are regarded as the behavioural classes. Clusters formed using clustering algorithms can be quite abstract. Intuitively, one tends to think of a behavioural class as, for instance, the ‘stressed traveller’ or the ‘experienced traveller’. However, the results of a cluster analysis are evidently not interpreted by the algorithm and are thus subject to human interpretation, although the abstractness of the results can make it difficult to put into words the meaning of a cluster.

3.2.1.3 Classification

The clustered data set forms the input for the classification part of the framework. The classification algorithm considers the passenger characteristics and the assigned behavioural class of each object in the data set. Based on this, it creates a classifier to assign objects to a specific behavioural class. With this classifier, new observations that were not part of the initial clustering data set can be assigned to a behavioural class.

Considering the presented framework from an operational point of view, it can be noted that the classification process is responsible for classifying newly observed passengers. Hence, this process is continuous; each newly observed passenger is passed through the classifier in order to assign this passenger to a class. The clustering process provides the behavioural classes and can be executed when desired. For instance, different clustering results may appear on weekends and weekdays and so it may be decided to perform a new clustering at the start of the week and at the start of the weekend.

3.2.2 Dataflow through the CC-framework

The previous subsection provided an overview of the clustering and classification process as a whole. This subsection further describes what data is used in the framework, and how it is used. Figure 3.2 illustrates this.

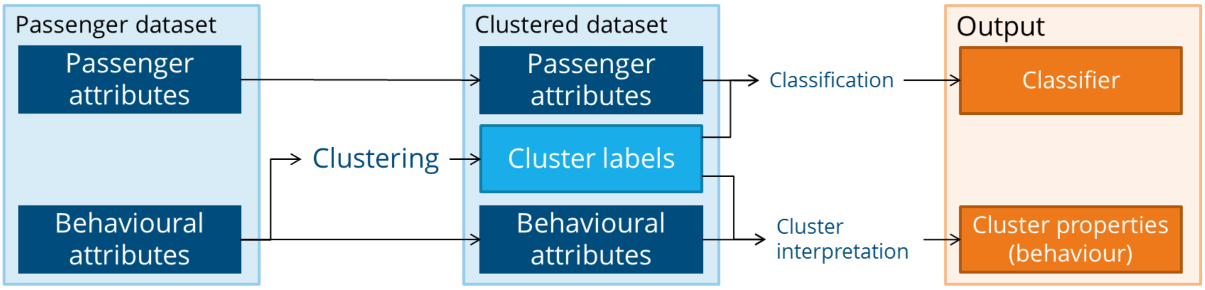


Figure 3.2: Dataflow through the framework

The leftmost blue block in the figure represents the passenger (or training) data set. This is a data set that contains a number of observations. Each observation represents one passenger and contains a number of attributes. These attributes are divided into two categories: passenger attributes and behavioural attributes. Behavioural attributes represent the behaviour of a passenger. This can pertain to any of the three basic levels of behaviour, as defined in section 2.4.2. Hence, the behavioural attributes represent the behaviour of the passenger and are consequently the attributes based on which behavioural groups have to be formed. Passenger attributes are the personal, process, and trip characteristics that describe the properties of passengers that are not regarded as behaviour.

The clustering part of the CC-framework finds the distinctive behavioural groups. Hence, the clustering technique is performed only based on the behavioural attributes in the observation data set. This yields the clustered data set, depicted in the middle of figure 3.2. This is the same data set as the observation data set, but augmented with the cluster label for each observation in the data set. This cluster label indicates to which of the clusters the observation has been assigned. Based on these cluster labels and the behavioural attributes in the data set, the properties of each cluster can be found. Hence, these cluster properties describe the behaviour of observations in the cluster. Furthermore, combining the cluster labels and the passenger attributes with a classification algorithm leads to the classification rules. These rules can assign new observations to a cluster solely based on their passenger attributes.

It can be seen in the figure that, although they are in the same data set, the passenger and behavioural attributes are used separately in the CC-framework. In summary: the behavioural attributes lead to the behavioural classes, while the passenger attributes lead to the rules to assign (new) observations to these behavioural classes.

3.3 Cluster analysis techniques

The first part of the CC-framework consists of clustering. Clustering, or cluster analysis, is a general term for machine learning methods that segment a set of data into groups, or clusters. With a clustering algorithm, a set of heterogeneous data is grouped into clusters in which each object is more similar to other objects within the cluster, than it is to objects in other clusters (Han, Pei, & Kamber, 2011). Consequently, clustering forms homogeneous groups of objects out of a heterogeneous set of objects. The clusters that are formed are not defined a priori, i.e., the algorithm determines the clusters. Depending upon the algorithm that was chosen, the number of clusters can be defined beforehand, or found during the clustering process. However, the interpretation of the meaning of these clusters does not follow from the clustering itself. This task is up to the researcher. Taking into account the CC-framework, there are several conditions a clustering technique should satisfy. Metrics to assess the performance with respect to accuracy of the clustering should be available. Because the passenger data will contain both categorical and numerical data, the technique should be able to handle mixed type data. Also, the technique should ideally be able to find the optimal number of clusters.

In this section, several clustering techniques are discussed and assessed. This includes conventional clustering techniques, as well as the model-based latent class analysis (LCA) technique. After this, the data input for clustering and how to assess the validity of clusters is discussed. The section then ends with a choice of a clustering technique. The implementation of the selected technique is discussed thereafter.

3.3.1 Conventional cluster analysis

Conventional clustering techniques separate objects based on their dissimilarity, expressed by some measure of distance. The conventional techniques can be segmented into two main types: hard clustering and soft clustering. In hard clustering, an object cannot be part of more than one cluster. In case of closely grouped objects that form distinct clusters, this can accurately represent the clusters in the data. Consider the example of figure 3.3, which represents a data set consisting of twelve objects. Using hard clustering, these objects have been clustered into two clusters. In the case of hard clustering, the red object falls in the right cluster. However, in this case, the red object may be better represented if it were not part of only one cluster, but a combination of the two. This can be done with soft – or fuzzy – clustering, which is presented in section 3.3.1.2.

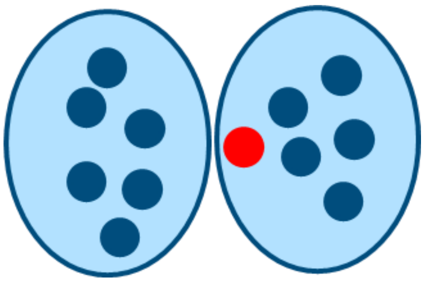


Figure 3.3: Hard clustering a data set in two clusters, the red object may be better represented by soft clustering

3.3.1.1 Hard clustering

In hard clustering, an object can belong to only a single cluster. Hard clustering algorithms can be divided into two main categories (Abonyi & Feil, 2007): hierarchical or partitional/non-hierarchical.

Hierarchical clustering

Hierarchical clustering yields a nested set of clusters. The complete data set is enclosed in an all-encompassing cluster, which is in turn iteratively divided into smaller, nested sub-clusters. This yields a number of nested clusters that can be represented by a dendrogram.

Hierarchical clustering is done by iteratively applying a clustering algorithm, which can be done in two ways. The first is a bottom-up approach, called agglomerative clustering. This method starts with every object in its own cluster and combines the most similar clusters in each step, until one large cluster is left. The second way of hierarchical clustering is a top-down approach, known as divisive clustering. This method starts with one cluster that contains all objects. During each iteration, the clusters are split to form new clusters that are less similar (Abonyi & Feil, 2007). The clustering process stops once a stopping criterion, such as a minimum or maximum number of objects in a cluster, is reached. This stopping criterion is mostly defined from a performance-wise point of view; the larger the number of levels in the dendrogram, the more computationally expensive the clustering is. Once the hierarchical clustering process is complete, the researcher has to choose the level of the dendrogram to use as clusters. Consequently, a smart choice has to be made, which can be based on for example the calculation of performance metrics to assess the fit of the clustering for each level of the dendrogram.

A fictitious example for a data set containing information about travellers is shown in figure 3.4. The corresponding dendrogram for this example is shown in figure 3.5. In the example, the traveller data set is clustered into two main clusters: cluster 1 and 2. Each of these two clusters is then divided into two sub-clusters.

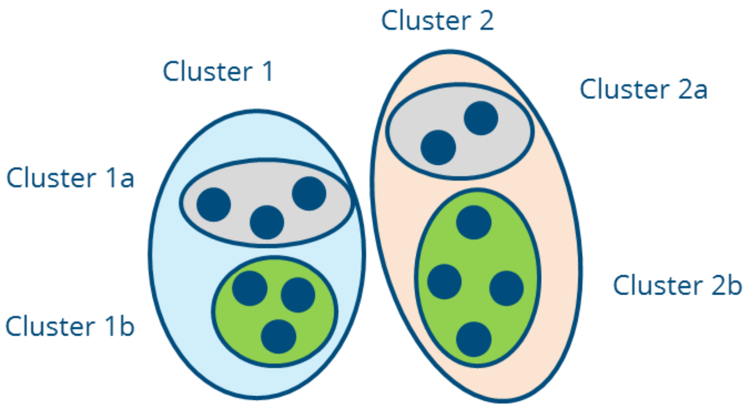


Figure 3.4: Hierarchical clustering of a traveller data set

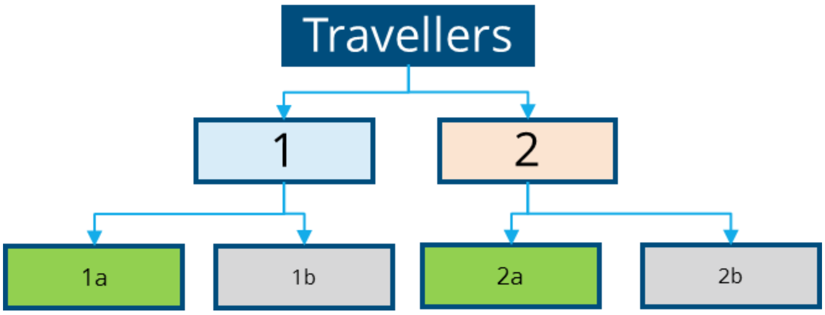


Figure 3.5: A dendrogram representing the hierarchical clustering of figure 3.4

Partitional/Non-hierarchical clustering

Partitional clustering yields a number of separate, non-overlapping clusters. Contrary to hierarchical clustering, there are no nested clusters, or sub-clusters. A partitional clustering can be seen as a 'slice' of one level of a hierarchical clustering. A main advantage of partitional clustering lies in the fact that it is much less computationally expensive than hierarchical clustering (Abonyi & Feil, 2007).

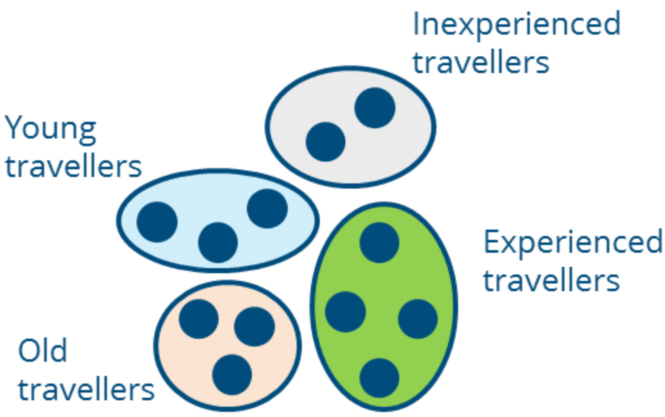


Figure 3.6: Partitional/Non-hierarchical clustering of a traveller data set

Figure 3.6 shows an example of a clustered data set about travellers. The clusters here are not part of a higher level cluster. Note that the cluster names serve only as an example.

3.3.1.2 Soft/Fuzzy clustering

In some cases, an object cannot be properly represented as a part of only one cluster. Fuzzy clustering, based on fuzzy logic, allows an object to be a member of all clusters. This membership is represented by a membership value that lies between 0 and 1 (inclusive). This is especially useful for clusters that are not clearly separated, or for data that has a lot of noise. For natural situations, fuzzy clustering can better represent the actual situation as objects are not forced to be part of one cluster, which may not fully describe the object's properties (Abonyi & Feil, 2007). Fuzzy clustering allows a combination of the properties of two or more clusters, which can better fit the object. Figure 3.7 illustrates the fuzzy clustering of a traveller data set. Note the two objects that are part of two clusters.

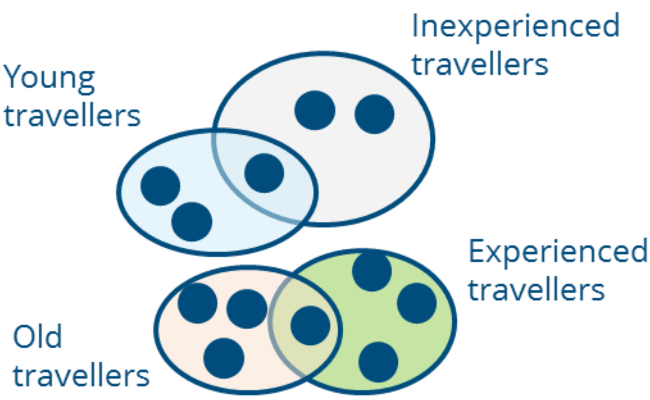
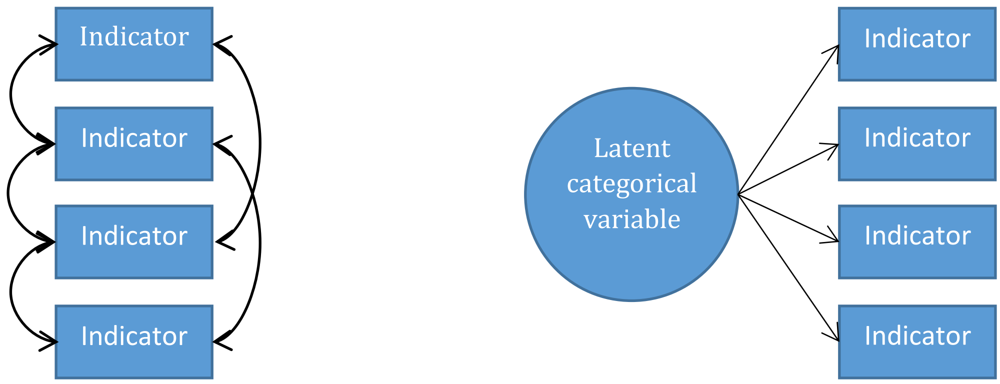


Figure 3.7: Fuzzy clustering of a traveller data set

Figure 3.7 presents a fuzzy clustering of the same data as the example in figure 3.6. Using hard clustering, an object could only be part of one cluster. In the case of this example, this means the traveller can either belong to the cluster 'Young', 'Old', 'Inexperienced', or 'Experienced'. Intuitively, this does not cohere with the natural situation. A traveller can be old, but also inexperienced. Conversely, a young traveller can also be experienced. This notion is solved using fuzzy clustering. An object can be part of any cluster, indicated by a certain membership value. Relating this to the example, a young, but experienced traveller can have a membership of 0.6 for the 'Young' cluster and a membership of 0.4 for the 'Experienced' cluster. This way, the traveller is represented by a combination of the properties of the two clusters. Depending on the chosen algorithm, this can also be a probabilistic interpretation.

3.3.2 Latent class analysis

Another method to find groups in data, related to the aforementioned types of clustering, is latent class analysis (LCA), also referred to as latent class cluster analysis (LCCA). Contrary to the aforementioned clustering methods, LCA relies on a statistical model to find groups in the data. The previously presented forms of cluster analysis mainly use the distance between objects to find groups in the data. As a consequence, the groups that are found in the data primarily follow from the dissimilarities in the data set. In contrast with these clustering methods, LCA is a model-based method that performs a probabilistic class assignment (Molin, Mokhtarian, & Kroesen, 2016). The main concept of LCA is that there exists an unknown latent variable that can account for the unobserved subgroups that are present in the data (Vermunt & Magidson, 2004). This latent variable entails the classes that are present in the data and accounts for the associations between the attributes in the data in such a way that these associations become insignificant. This important aspect of LCA is called the assumption of local independence (Vermunt & Magidson, 2004). The principle of LCA is visualized in figure 3.8. Figure 3.8a represents the observed variables, called indicators, which have some interrelation. In figure 3.8b, these relations are accounted for by the latent variable. Hence, there are not mutual association between the indicators anymore. Latent class analysis has been proven to yield good results and even significantly better results compared to, for example, k-means clustering[1] (Magidson & Vermunt, 2002).



(a)
Without LCA, there are associations between the observed indicators

(b)
With LCA the latent variable accounts for these associations and contains the classes

Figure 3.8: Principle of LCA, adapted from Molin et al. (2016)

3.3.3 Data input

The aforementioned types of clustering methods require input data that is to be clustered. Such data can be characterised as follows. The input data contains N objects to be clustered. An object has Y features, also called attributes. Relating this to the case of an airport terminal, an object is a passenger. Features of such an object can include characteristics of the passenger, for example age or sex, but also information about his trip, such as Schengen or non-Schengen. Similar to the two types of characteristics: passenger and behavioural, as defined in section 2.2.1, the attributes in the data set can also be characterised as passenger attributes and behavioural attributes. Again, behavioural attributes are used to form behavioural classes during clustering, while passenger attributes are used during classification to assign passengers to a class.

The complete set of N objects and n features can be represented as an $N \times Y$ matrix $\boldsymbol{X}$ that contains the data that is to be clustered:

[1] K-means clustering is a well-known partitional hard clustering algorithm

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,Y} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,Y} \end{bmatrix} \tag{3.1}$$

3.3.3.1 Conventional cluster analysis

Conventional cluster analysis is based on the similarity and dissimilarity of objects; the distance between objects is hence important information for clustering algorithms. For some clustering algorithms, such as hierarchical algorithms, the distance between objects is the main input (Abonyi & Feil, 2007). These distances are represented in the form of a dissimilarity matrix:

$$\begin{bmatrix} 0 & d(1,2) & d(1,3) & d(1,Y) \\ & 0 & d(2,3) & d(2,Y) \\ & & 0 & d(3,Y) \\ & & & 0 \end{bmatrix} \tag{3.2}$$

In the dissimilarity matrix, $d(i,j)$ represents the distance measure between two objects i and j. Note that $d(i,j) = d(j,i)$ and hence the dissimilarity matrix is symmetrical. Additionally, the distance between an object and the object itself is obviously zero, hence: $d(i,i) = 0, \forall i$.

The distance measure in the dissimilarity matrix can be calculated in different ways and the chosen distance measure affects the outcome of the clustering algorithm. The distance measure should hence be chosen wisely (Abonyi & Feil, 2007). The best known example of a distance measure is the Euclidean distance, which is similar to the Euclidean norm of a vector. For a pair of two objects i and j, the Euclidean distance can be calculated as in (3.3).

$$d(i,j) = \sqrt[2]{\left(x_{i,1} - x_{j,1}\right)^2 + \left(x_{i,2} - x_{j,2}\right)^2 + \cdots + \left(x_{i,n} - x_{j,n}\right)^2} \tag{3.3}$$

Other well-known distance measures include the Minkowski distance, Manhattan distance and Mahalanobis distance (Abonyi & Feil, 2007; Xu & Wunsch, 2005). However, these distance measures, including the Euclidean distance, are only valid for numerical attributes. Additionally, data sets that have various numerical attributes with different scales should be normalised so as to not skew the distance measure due to differences in magnitude of the numerical values.

Data sets that contain mixed attributes require a distance measure that is also able to handle categorical attributes. This is necessary for the present case as passenger behaviour can be represented by mixed attributes, for example: a numerical attribute for the time spent in the lounge and a categorical attribute that describes whether or not the passenger has visited any shops. One possibility to tackle this problem is to the categories of a categorical variable into several dummy variables. However, a special distance measure is also possible, such as the Gower distance (Gower, 1971). The Gower distance calculates the normalised distance between numerical attributes:

$$d^a(i,j) = \frac{\left|x_i^a - x_j^a\right|}{\max(\boldsymbol{x}^a) - \min(\boldsymbol{x}^a)} \tag{3.4}$$

Categorical attributes are compared and can either be given a distance of *1* or *0*, dictated according to (3.5):

$$\begin{cases} d^a(i,j) = 0, & x_i^a = x_j^a \\ d^a(i,j) = 1, & x_i^a \neq x_j^a \end{cases} \tag{3.5}$$

The distance measure also accounts for a special case of binary attributes, where the presence of an attribute in both compared objects causes a higher similarity than the absence of an attribute in both objects. The total Gower distance is calculated with the mean of the distance measures for numerical, categorical and binary attributes.

3.3.3.2 Latent class analysis

In contrast with conventional cluster analysis, LCA is not based on dissimilarity between objects. Rather than this, it is based on a probabilistic model based on the classes and the outcomes of the attributes in the data. In short, this model optimizes the class-conditional outcome probabilities of attributes and the class membership probabilities of objects. As such, LCA does not rely on the distance between objects as conventional cluster analysis does. It is therefore not necessary to pick a fitting distance measure for LCA.

3.3.4 Validity of results

Clusters that are formed during cluster analysis can be difficult to validate. In case of well-segmented data sets that have two or three attributes, it is possible to inspect the clustering results in 2D or 3D space, respectively. An example of a simple clustering of three clusters with two attributes, hence represented in a two dimensional graph, is shown in figure 3.9. It is evident that resulting clusters (red, green and blue) correctly represent the three groups that are present in the input data.

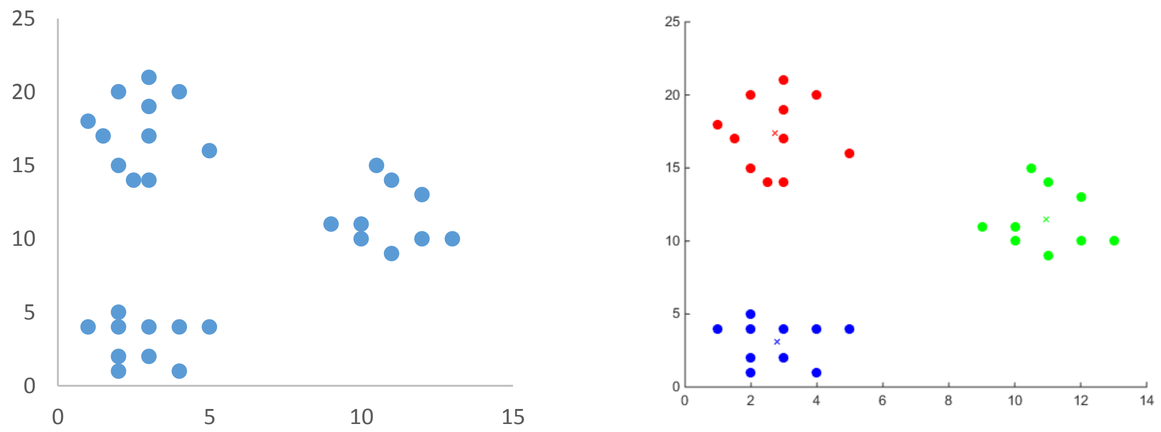


Figure 3.9: Simple clustering example, left: input, right: clustered output

However, data sets that contain more than three attributes, or that contain categorical attributes, can be less easily represented visually. Moreover, even if they are visually represented, groups present in the data may not be visually discernible. Therefore, validity measures are necessary to assess the correctness of the achieved clustering results.

3.3.4.1 Conventional cluster analysis

There are various options available to assess the clustering validity of conventional clustering techniques. Two cases can be distinguished, depending on the availability of the so called *ground truth* (Han et al., 2011). This ground truth contains the actual class labels which are regarded as the perfect clustering. Obviously, the ground truth is only available for training data sets. If the ground truth is available, extrinsic validity methods can be used. However, in the case of behavioural classification, the ground truth is unknown. In such case, intrinsic validity methods can be used. For the interested reader, a short overview of extrinsic and intrinsic indices is given in Appendix B.

3.3.4.2 LCA

Because LCA is a model-based statistical method, it has its own performance metrics that represent the relative goodness of fit of the model. The two most used metrics are the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) that represent the goodness of fit of the model, adjusted for the complexity of the model. This allows for finding the best fitting number of classes for the data.

3.3.5 Choice of method and implementation

The previous sections have introduced a number of techniques that can be used to find groups in data. Cluster analysis finds groups in data based on the dissimilarity of objects in the data. This brings along the difficulty of defining the distance between objects, certainly in the case of mixed variable type data. Because many algorithms are aimed at clustering continuous data, clustering mixed data requires reworking the used distance measure. Latent class analysis, originally designed for dichotomous variables only, is better able to handle categorical variables.

Additionally, LCA has been shown to yield better results compared to some algorithms (Magidson & Vermunt, 2002). To test this, both conventional cluster analysis and LCA have been tested on mixed variable type test data sets from the UCI Machine Learning Repository (Lichman, 2013), which is comparable to the type of data that would form the input for the CC-framework. Indeed, LCA performs better on these data. Detailed results, comparing LCA and conventional cluster analysis methods, are included in Appendix B. Additionally, because LCA is model-based, it offers performance measures that allow for easily comparing models. For these reasons, it was decided to further use LCA in this thesis work. In spite of this, some work has also been put in implementing the other clustering algorithms. Although these did not yield particularly good results, some information for the interested reader is included in Appendix B.

Hereafter, some more information about the specific workings of LCA is introduced. Additionally the implementation is discussed along with the performance metrics that follow from it and how to interpret these.

3.3.5.1 Clustering using Latent Class Analysis

Latent class analysis can be implemented in a number of ways. There are dedicated software packages available, such as Latent GOLD®. However, keeping in mind our goal of implementing a sequential clustering and classification method, such a separate piece of software would not be practical as it introduces additional manual effort between the clustering and classification steps. Therefore it was decided to implement LCA using the R programming language (R Core Team, 2015). Although mainly aimed at statistical applications, R offers plenty flexibility as a programming language and benefits from a high number of available packages.

The LCA was implemented in R using the 'poLCA' package (Linzer & Lewis, 2011). poLCA is able to estimate a latent class model for polytomous, i.e. categorical, variables. As a consequence, any continuous variables included in the model have to be binned first. For a full description of the algorithm used in poLCA, the reader is referred to the work of Linzer and Lewis (2011). Summarizing their work, the steps in poLCA can be described as follows:

- Define the following:
    - There are J categorical, observed outcome variables, called manifest variables.
    - Each of these variables can have K_j possible outcomes.
    - There are N objects in the data.
    - Y_{ijk} denotes the observed values of the manifest variables, here:
        - $Y_{ijk} = 1$ if the object i has the kth outcome for the jth variable
        - $Y_{ijk} = 0$, otherwise
    - R denotes the number of classes to be estimated
    - π_{jrk} denotes the probability that an observation in class r produces the j^{th} outcome for the k^{th} manifest variable.
- The probability density function for all classes is defined as:
    - $P(Y_i|\pi, p) = \sum_{r=1}^{R} p_r \prod_{j=1}^{J} \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}}$
- p_r and π_{jrk} have to be estimated by the latent class model, this is done by maximizing the following log-likelihood function:
    - $\ln L = \sum_{i=1}^{N} \ln P(Y_i|\pi, p)$

- Optimise the log-likelihood function is using the Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977), beginning with arbitrary values for p_r and π_{jrk}.
- Calculate and output the posterior class-probabilities per object, which indicate the likelihood of belonging to each class.

3.3.5.2 Performance metrics

As mentioned before, a major benefit of LCA is that it has many goodness of fit indicators available that can help choosing the best-fitting model. The log-likelihood, as calculated during the estimation of the model can indicate model fit. However, this does not account for the complexity of the model and hence overfitting the data. As such, the log-likelihood always improves as the number of classes (and hence the complexity) in the model increases, while essentially the model is overfit. Fortunately, there are indicators available that compensate for this risk.

poLCA outputs the two most widely used of these indicators. Both of these indicators are based on the maximum log-likelihood of the model and a correction based on the number of estimated parameters in the model (Linzer & Lewis, 2011).

The first indicator is the Akaike information criterion (AIC), as introduced by Akaike (1998):

$$AIC = -2LL + 2\phi \tag{3.6}$$

Here, LL is the maximum log-likelihood of the estimated model and ϕ the number of estimated parameters. Essentially, this is negative two times the maximum log-likelihood, to which two times the number of estimated parameters has been added. This way, overly complex models (with a lot of parameters) are penalized.

The second indicator is the Bayesian information criterion (BIC), as introduced by Schwarz (1978):

$$BIC = -2LL + \phi \ln N \tag{3.7}$$

The BIC is essentially an adapted version of the AIC, which also takes into account the number of objects in the model, indicated by N. This way, also the statistical goodness of fit of the model is taken into account.

Both the BIC and AIC indicate a better fit of the model for a lower value of the indicator. However, they are both relative and can only be used to compare between models; there is no threshold value indicating a good model. Nevertheless, the model fit can be tested to compare the metrics between a model with only one class and a model with two or more classes. Lower values for the BIC and AIC for the latter case would indicate that indeed an LCA with more than one class is appropriate for the data.

Though both metrics often agree on the best model, the AIC has the risk of preferring an overfit, while the BIC may favour an underfit (Dziak, Coffman, Lanza, & Li, 2012). Additionally, the BIC is consistent, while the AIC is not. This means that the BIC is able to identify the smallest adequate model, while the AIC retains the risk of selecting an overly complex model as the number of objects N becomes too large (Dziak et al., 2012).

3.4 Classification methods

The previously discussed clustering algorithms find classes in data and label the data accordingly. These groups are based on the behavioural attributes in the data set and are hence behavioural classes. Now, as a passenger enters the terminal, only passenger attributes of this passenger can be collected. Consequently, based on these passenger attributes, it should be predicted to which behavioural class the passenger belongs. This is where classification methods, which form the second part of the CC-framework, come into play.

Classification methods predict an object's class based on its attributes. Hence, in the present case, the classification method should predict the behavioural class based on the passenger attributes of a passenger. Preferably, the classification method should be transparent so that it is clear how the classifier performs classification. Additionally, the method is preferably quick to compute, accurate, and able to deal with both continuous and categorical attributes.

There are many ways to create a classifier to perform such a classification. In the subsequent subsections, the overall classification process and various types of classification algorithms are discussed. Additionally, the performance measures of the classifier are discussed. Finally, a classification method is chosen and the implementation of the classification method is presented.

3.4.1 Classification process

Generally, data classification can be described as a two-step process (Han et al., 2011). An overview of this process is given in figure 3.10. The first block represents the data for which a classifier has to be constructed. These data contain a number of objects with various attributes and a class label. The classification algorithm finds the rules[2] that relate the attributes to a certain class. In the case of this thesis, the class labels are the clusters as they have been found during cluster analysis, as described in section 3.3.

When constructing a classifier, there is always a risk of overfitting the data. This means that the classifier is based too much on the specific properties and quirks of the data set it was made with. An overfitted classifier would hence perform well on the data set it was made with, but not with other data. To be able to check for, and prevent overfitting, the data set is split into a training subset and a test subset. The first is used to create the classifier, while the latter is used to assess the performance of the classifier. There are multiple methods to create the training and test data sets. A very common method is the holdout method, which randomly picks a number of objects for the training data set, leaving the remaining objects for the test data set. A common partitioning is 2/3 of the data for the training set and 1/3 for the test set (Han et al., 2011; Kotsiantis, Zaharakis, & Pintelas, 2007).

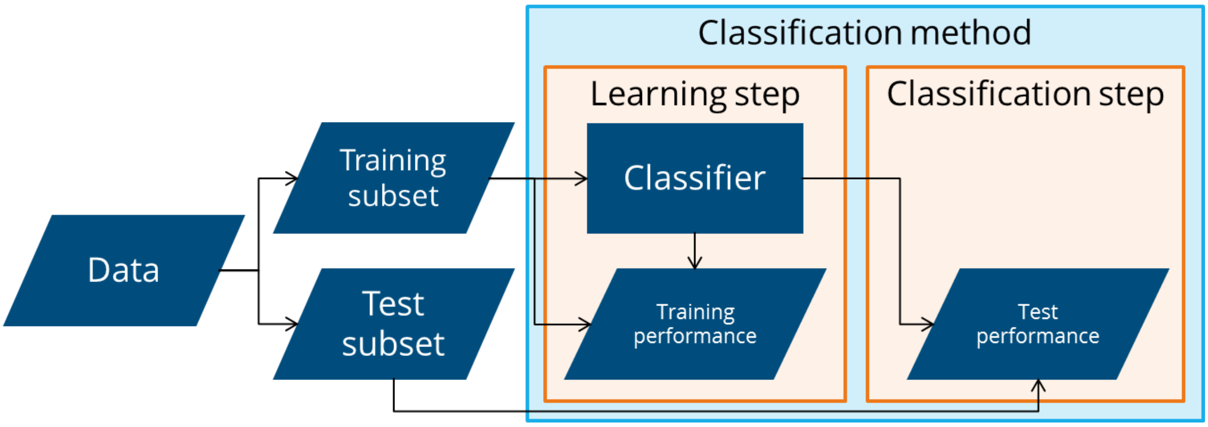


Figure 3.10: Classification algorithm process

The training data set is then used for the first step in the classification process: the learning step. Based on the attributes and classes of the objects in the training data set, a classifier is constructed. This classifier contains some form of rules that can predict the class an object belongs to, based on its attributes. The accuracy of this classifier can be calculated based on the classes predicted by the classifier and the data's actual classes. In the next step, which is the classification step, 'new' objects from the test set are classified by the classifier. Because this test data was not used when

[2] 'Rules' is used here as a general term for the way a classifier assigns objects to a class. While indeed classification can be in the form of logical rules (e.g.: IF X > Y THEN class A), this is not the case for all algorithms.

constructing the classifier, the performance on the test data gives a more accurate reflection of the actual performance of the classifier.

3.4.2 Types of classification algorithms

There is a multitude of classification algorithms available and described in literature. Five main categories of classification algorithms can be distinguished (Kotsiantis et al., 2007), of which a very concise overview is given below:

- **Logic based algorithms** provide logical rules to classify objects based on their attribute values.
- **Perceptron-based techniques** combine the weighted attribute values of the object to be classified and assign the object to a class based on a certain threshold.
- **Statistical learning algorithms**, rather than deterministic models, are based on probability models to predict the probability that an object belongs to a class.
- **Instance-based learning algorithms**, also known as lazy-learning algorithms, do not construct a classifier beforehand. Instead, these algorithms classify every new instance separately based on objects in the training data that are similar to the input that is given to the algorithm.
- **Support vector machines** perform binary classification by constructing a hyperplane that separates the two classes.

Although some of the mentioned algorithms, such as support vector machines, are designed for binary classification, these can be extended to support multiclass classification (Han et al., 2011). This is done by combining multiple binary classifiers that each classify in a one-versus-many fashion. Each of the aforementioned classifiers has its own benefits with regards to accuracy, speed of classifier construction, speed of classification, et cetera. Additionally, not all the algorithms are suitable for either continuous or discrete variables (Kotsiantis et al., 2007).

The aforementioned algorithms can be used as-is. However, classification accuracy can be improved by using ensemble methods (Han et al., 2011). Ensemble methods combine multiple, often relatively simple, classifiers to form one ensemble classifier. The classification made by the ensemble is a result of the weighted vote of each classifier in the ensemble. Two major forms of ensemble methods are boosting and bagging:

- **Boosting** creates classifiers for the training data many times, but focusses on misclassified objects. After each iteration, misclassified objects are identified and assigned a higher weight compared to correctly classified objects. These weights are taken into account when constructing a classifier in the next iteration, hence increasing the accuracy for these misclassified objects. The final, boosted classifier consists of the combination of the weighted votes of all individual classifiers (Alfaro, Gámez, & Garcia, 2013; Han et al., 2011).
- **Bagging** (bootstrap aggregation) also creates many classifiers based on the training data. However, the training data set is bootstrapped[3] before each iteration. The bagged classifier's classification is based on the majority vote of the individual classifiers in the ensemble (Alfaro et al., 2013; Han et al., 2011).

3.4.3 Evaluating classifier performance

The performance of a classifier can be assessed by comparing the class labels that were predicted by the classifier to the actual class labels of the data. There are at least 24 different metrics available for various kinds of classification problems (Sokolova & Lapalme, 2009). None of these metrics is all-encompassing; hence a few different metrics should be used to assess a classifier's performance. Consequently, several metrics have been selected and will subsequently be introduced. As main

[3] During bootstrapping, the data set is sampled with replacement for the same number of objects as are in the data set. Hence, the data set remains the same size, but some original elements will appear multiple times, while others will be missing completely. This can help reduce noise in the data set.

sources for this, the author has thankfully used the work of Han et al. (2011) and Sokolova and Lapalme (2009).

3.4.3.1 Confusion matrices

In order to explain the different metrics, some basic terminology regarding classification performance needs to be established. The basis for all metrics is the confusion matrix, which compares the predicted class labels, as predicted by the classifier, to the actual class labels. For a classification problem with two classes, i.e. binary classification, the confusion matrix can be represented as in table 3.1.

Table 3.1: Binary confusion matrix

		Predicted class	
		Positive	**Negative**
Actual class	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

The cells in the binary confusion matrix contain the number of objects that have been classified as such. The binary confusion matrix can be extended to multiclass problems by extending the number of rows and columns. However, from such a multiclass confusion matrix the number of true positives and true negative etcetera is not directly visible. To obtain the TP, FN, FP and TN values for a multiclass problem, one-versus-all binary confusion matrices can be constructed for each class. Consider the example classification presented in table 3.2, containing nine objects and three classes.

Table 3.2: Example of a multiclass classification

Object #	Actual class	Predicted class
1	1	1
2	1	3
3	1	1
4	2	3
5	2	2
6	2	2
7	3	1
8	3	1
9	3	3

A multiclass confusion matrix can be constructed for this classification, leading to the result in table 3.3. The numbers on the diagonal of this matrix represent the true positives for class 1, 2 and 3. Let us call this matrix C and indicate the elements of this matrix as $c_{i,j}$. Now, looking at class 1, the number of true negatives can be found as the sum of the elements $c_{2,2}, c_{2,3}, c_{3,2}, c_{3,3}$. These elements are the objects that do not belong to class 1 and are also not classified as such. The number of false positives is found as the sum of elements $c_{2,1}$ and $c_{3,1}$, which are the objects that do not actually belong to class 1, but have been classified as such. Lastly, the number of false negatives for class 1 is found as the sum of elements $c_{1,2}$ and $c_{1,3}$. These are the objects that do actually belong to class 1, but have not been classified accordingly.

Table 3.3: Multiclass confusion matrix

		Predicted class		
		1	**2**	**3**
Actual class	**1**	2	0	1
	2	0	2	1
	3	2	0	1

Completing this for all three classes yields three separate one-versus-all binary classification matrices. These matrices are shown in table 3.4.

Table 3.4: Binary confusion matrices for a multiclass classification

	Class 1	Predicted	
		Pos	Neg
Actual	Pos	2	1
	Neg	2	4

	Class 2	Predicted	
		Pos	Neg
Actual	Pos	2	1
	Neg	0	6

	Class 3	Predicted	
		Pos	Neg
Actual	Pos	1	2
	Neg	2	4

3.4.3.2 Accuracy and error

The two most basic metrics are the accuracy and error. The accuracy represents the fraction of objects that was correctly classified, whereas the error indicates the fraction of objects that was classified incorrectly. Hence, accuracy and error are complementary.

However, these two metrics are defined differently for binary and multiclass problems. For binary classification the accuracy can be calculated as in (3.8). The error is calculated as in (3.9).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.8}$$

$$Error = \frac{FP + FN}{TP + TN + FP + FN} = 1 - Accuracy \tag{3.9}$$

For a multiclass classification, the average accuracy is used (Sokolova & Lapalme, 2009). The complement of the average accuracy is called 'error rate', but in order to differentiate from the binary case, the term 'average error' will be used here. The average accuracy and average error are shown in (3.10) and (3.11). Here, C indicates the number of classes. Hence, the average accuracy is the average of the accuracies as calculated for the C one-versus-all binary confusion matrices. The average error is calculated is a similar fashion.

$$Average\ Accuracy = \frac{\sum_{c=1}^{C} \frac{TP_c + TN_c}{TP_c + TN_c + FP_c + FN_c}}{C} \tag{3.10}$$

$$Average\ Error = \frac{\sum_{c=1}^{C} \frac{FP_c + FN_c}{TP_c + TN_c + FP_c + FN_c}}{C} \tag{3.11}$$

In addition to the average accuracy and average errors, some authors tend to use another metric as a representation of accuracy for a multiclass classification (Sokolova & Lapalme, 2009), such as Alfaro et al. (2013) and S. Zhu, Ji, Xu, and Gong (2005). In this definition, accuracy is simply defined as the total number of correctly classified objects, divided by the total number of objects. To differentiate this metric from the others, in this work it is defined as in (3.12) and (3.13). Where CC is the number of correctly classified objects, IC the number of incorrectly classified objects, and N the total number of objects in the data set.

$$Overall\ Accuracy = \frac{CC}{N} \tag{3.12}$$

$$Overall\ Error = \frac{IC}{N} = 1 - accuracy \tag{3.13}$$

To illustrate the difference between these various metrics for accuracy (and error), these values have been calculated for the example multiclass classification of section 3.4.3.1. Table 3.5 shows the resulting values. The 'Accuracy' column shows the per-class one-versus-all accuracy. The average accuracy is simply the average of these values. By contrast, the overall accuracy is the sum of the diagonal of the multiclass confusion matrix (table 3.3), divided by the total number of objects.

Table 3.5: Per class one-vs-all accuracy, average accuracy and global accuracy for the example of Table 3.2

Class	Accuracy	Average Accuracy	Overall Accuracy
1	${}^{6}/_{9} \approx 0.67$	$\frac{\left({}^{6}/_{9} + {}^{8}/_{9} + {}^{5}/_{9}\right)}{3} \approx 0.70$	$\frac{2+2+1}{9} \approx 0.56$
2	${}^{8}/_{9} \approx 0.89$		
3	${}^{5}/_{9} \approx 0.56$		

3.4.3.3 Precision, recall and F-score

The average accuracy and overall accuracy are very intuitive and easily interpretable metrics. However, they suffer from the class imbalance problem (Han et al., 2011). This means that these metrics work well when the distribution of objects over the classes is uniform, but will not properly represent accuracy when the classes are imbalanced. To illustrate this, consider an example with three classes. Assume that 98 objects are of class 1, while classes 2 and 3 have one object each. Assume that a classifier classifies all objects into class 1. The overall accuracy of this classifier would be 98% and the average accuracy 97%. It goes without saying that the actual performance of the classifier is thus not reflected by these accuracy metrics.

The precision and recall metrics can be used to overcome this class balance problem (Han et al., 2011). The precision measure represents the fraction of objects that have been correctly classified to a class out of the total number of objects that have been classified to that class. Conversely, recall represents the fraction of correctly classified objects out of the total number of objects actually in that class. The two metrics can either be defined as micro-averaged, such as in (3.14) and (3.15), or as macro averaged, such as in (3.16) and (3.17). Whereas the micro-averaged precision and recall favour bigger classes, the macro-averaged precision and recall treat all classes equally (Sokolova & Lapalme, 2009).

$$Micro\ Precision = \frac{\sum_{c=1}^{C} TP_c}{\sum_{c=1}^{C} (TP_c + FP_c)} \tag{3.14}$$

$$Micro\ Recall = \frac{\sum_{c=1}^{C} TP_c}{\sum_{c=1}^{C} (TP_c + FN_c)} \tag{3.15}$$

$$Macro\ Precision = \frac{\sum_{c=1}^{C} \frac{TP_c}{TP_c + FP_c}}{C} \tag{3.16}$$

$$Macro\ Recall = \frac{\sum_{c=1}^{C} \frac{TP_c}{TP_c + FN_c}}{C} \tag{3.17}$$

To illustrate the introduced metrics, these have been calculated for the example classification of table 3.2. The results are shown in table 3.6. It can be seen that the classifier performs best for class 2 and worst for class 3. Overall, the macro-averaged precision is a bit more optimistic compared to the micro-averaged precision. This is because the class 1 and 3, for which the classifier performs worst, have more objects classified to them compared to the better performing class 2.

Table 3.6: Precision and recall for the example of table 3.2

Class	Precision	Recall	Micro Precision	Micro Recall	Macro Precision	Macro Recall
1	$\frac{2}{2+2} = 0.5$	$\frac{2}{2+1} \approx 0.67$	$\frac{2+2+1}{2+2+2+0+1+2} \approx 0.56$	$\frac{2+2+1}{2+1+2+1+1+2} \approx 0.56$	$\frac{\frac{2}{4}+\frac{2}{2}+\frac{1}{3}}{3} \approx 0.61$	$\frac{\frac{2}{3}+\frac{2}{3}+\frac{1}{3}}{3} \approx 0.56$
2	$\frac{2}{2+0} = 1.0$	$\frac{2}{2+1} \approx 0.67$				
3	$\frac{1}{1+2} \approx 0.33$	$\frac{1}{1+2} \approx 0.33$				

Precision and recall both indicate different aspects of a classifier's performance. Consequently, a related metric that combines the two is the F-score, see (3.18). The F-score is based on the harmonic mean of the precision and recall. It can be computed either with the micro-averaged precision and recall, or with the macro-averaged precision and recall. β is a non-negative real number, for which commonly chosen values are 2 and 0.5 (Han et al., 2011).

$$Fscore = \frac{(\beta^2 + 1) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \tag{3.18}$$

3.4.3.4 ROC curves and AUC

In addition to the aforementioned metrics, Receiver Operating Characteristic (ROC) curves can visualise the trade-off between the sensitivity and specificity (Han et al., 2011), shown in (3.19) and (3.20). The sensitivity, which is in fact the same as recall, shows the amount of positives that have been correctly classified as such. The specificity is the amount of negatives that have been correctly classified as such. As opposed to the previously introduced metrics, ROC curves are not sensitive to changes in the class distribution (Fawcett, 2006).

$$Sensitivity = \frac{TP}{TP + FN} \tag{3.19}$$

$$Specificity = \frac{TN}{FP + TN} \tag{3.20}$$

A detailed explanation of ROC curves would be too elaborate for this report. Despite this, some basic notions about ROC curves have to be made. An ROC curve plots the sensitivity against 1-specificity (essentially, these are the true positive rate and the false positive rate). Evidently, as the number of true positives rises, the number of false positives also rises. For a good model, the false positive rate rises less than the true positive rate.

The class probabilities as they have been predicted by the classifier are needed to construct the curve. Based on the class probability, it can be decided if an object belongs to a class. For example, an object that has a class probability of 90% is very likely to belong to that class. However, for an object with a probability of 50%, this is not so certain. When constructing an ROC curve the threshold for the class probability is varied. The curve starts with a very high cut-off (low false positive rate) and ends with a very low cut-off (high false positive rate), plotting the sensitivity versus 1-specificity on the curve.

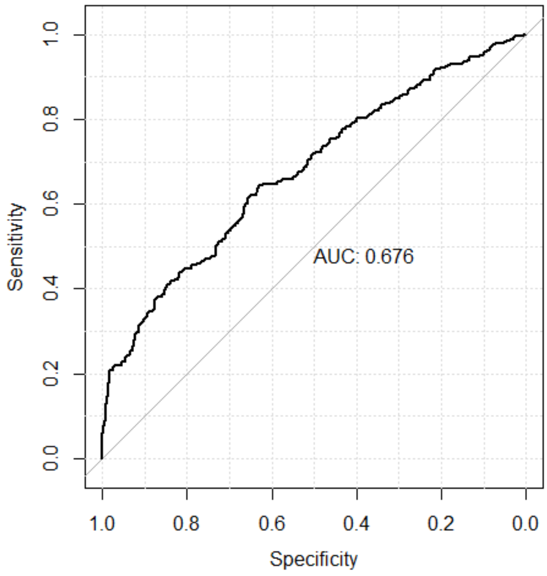


Figure 3.11: Example of an ROC curve

Figure 3.11 shows an example of an ROC curve. As mentioned, ideally the true positive rate is high, while the false positive rate is low. Hence, the ROC curve for a good classifier will come very close to the top left corner of the graph. The diagonal line represents the situation where the true positive rate equals the false positive rate. Essentially, this represents random guessing. Based on this notion, the performance of a classifier is often expressed as the Area Under the Curve (AUC). As the area under the diagonal is 0.5, the performance of the classifier compared to random guessing can be assessed. The AUC can also be used to compare different classifiers.

The ROC curve and the AUC are in principle defined for binary classification problems. However, there are methods to create these for multiclass classification problems. One possibility is to create a separate ROC curve for each class in a one-versus-all fashion (Fawcett, 2006). The AUC can then be calculated as a weighted average of the AUCs of the individual ROC curves.

3.4.4 Choice of method and implementation

The previous few sections have introduced a number of classification algorithms and metrics to assess their performance. In this section, an algorithm is chosen and the implementation is explained in further detail. The implementation has been done using the R programming language (R Core Team, 2015).

3.4.4.1 Classification

Based on their theoretical benefits over other classification methods, decision trees have been chosen. This logic-based method has several benefits that make it very suitable for our application. First, decision trees are very transparent with respect to how a classification is established. For a single decision tree, it is very simple to retrace the steps that the classifier has taken to classify an object to a certain class. Second, decision trees are relatively quick to compute (Kotsiantis et al., 2007). Third, decision trees are able to deal with both continuous and discrete values. This makes them very flexible with respect to the type of input variables.

Unfortunately, the general accuracy of decision trees is worse than some other algorithms (Kotsiantis et al., 2007), such as the currently very popular neural networks. Therefore, boosting is used as this has been shown to yield a highly accurate classifier (Alfaro et al., 2013; Dietterich, 2000; Opitz & Maclin, 1999). Boosting has been chosen instead of bagging as the former is primarily aimed at increasing accuracy, while the latter is primarily aimed at reducing variance. However, because boosting leads to many decision trees for one classifier, this decreases the interpretability of the classifier in comparison with a single decision tree. However, this is regarded as an acceptable price for increased accuracy.

This boosted decision tree approach has been implemented using the 'adabag' package for R (Alfaro et al., 2013). The specific algorithm employed is the 'Stagewise Additive Modeling using a Multi-class Exponential loss function' (SAMME) algorithm, introduced by J. Zhu, Rosset, Zou, and Hastie (2006). The algorithm is an adaptation of the AdaBoost algorithm (Freund & Schapire, 1996), which is the best known and well performing boosting algorithm, but only suitable for binary classification. SAMME is also suitable for multiclass classification and does so with good results (J. Zhu et al., 2006).

For a full description of the algorithm, the reader is referred to the work of J. Zhu et al. (2006), where the algorithm is first introduced. On a high level, the SAMME algorithm consists of the following steps:

1. Define the total number of decision trees to construct as B, chosen by the user.
2. For a data set N containing n objects, the observation weights are initialised as $w_b(i) = \frac{1}{n}$ where $i \epsilon N$. These observation weights indicate the importance of each object when fitting a decision tree.
3. For $b = 1, 2, \dots, B$, where B is the total number of trees to use in the boosting
    a. Fit a classifier to the data set using weights w_b
    In this case, decision trees are used as the classifier.
    b. Compute the error, which is weighted based on the weights w_b
    c. Compute the constant α_b, which is used to update the weights w_b
    This constant is used to check if the classifier result is better than random guessing.
    d. Update the weights for the next iteration w_{b+1} and normalise them.
    e. Return to step 2b and construct the next tree using the updated weights.
4. Output the final ensemble classifier. This classifier calculates the sum of the weighted votes of the decision trees in the classifier. The class with the highest vote is assigned to a classified object.

3.4.4.2 Performance metrics

The adabag package that is used to create the ensemble classifier is only able to output the classification accuracy. This is rather limited when assessing classifier performance. Additionally, the error is always calculated in the same way, regardless of the number of classes. In the adabag package, the error is calculated as $1 - \frac{CC}{N}$. Here, CC is the number of correctly classified objects and N is the total number of classified objects. While this is in fact correct for a binary classification, it does not represent the average accuracy as would be used in for a multiclass classification. Rather, it represents the exact match ratio, or overall accuracy, as defined in section 3.4.3.2. To overcome this confusion and extend the number of available metrics, the author has created a function to calculate the metrics of sections 3.4.3.2 and 3.4.3.3. That is: the average accuracy, overall accuracy, micro- and macro-averaged precision and recall, and the F-score. The implementation is as described in sections 3.4.3.2 and 3.4.3.3, consisting of the creation of a confusion matrix and, if necessary, multiple one-versus-all matrices.

The calculation of (multiclass) ROC-curves and the AUC is somewhat more involved. Fortunately, existing R packages that can do this are available. For the implementation of these metrics, the R package 'pROC' was used, created by Robin et al. (2011). The pROC package is also able to calculate the multiclass AUC. This calculation is based on the pairwise comparison of classes, and is implemented based on the work of Hand and Till (2001).

Some authors present the AUC as preferable over the other performance metrics (Bradley, 1997), such as average accuracy. While it is overall a widely used single-number metric (Hand & Till, 2001), other authors argue that it can also be misleading (Lobo, Jiménez-Valverde, & Real, 2008). Therefore, many different performance metrics have been implemented. In the ensuing, the results of all these metrics will be considered. The F-score will be used primarily as it combines two important aspects of a classifier's performance: precision and recall.

3.5 Integrated implementation

Sections 3.3.5 and 3.4.4 have presented the implementation of the clustering and classification methods. The combination of these two components forms the actual implementation of the clustering and classification framework as presented in section 3.2. Figure 3.12 gives a schematic overview of the implementation, which will be further discussed in this section.

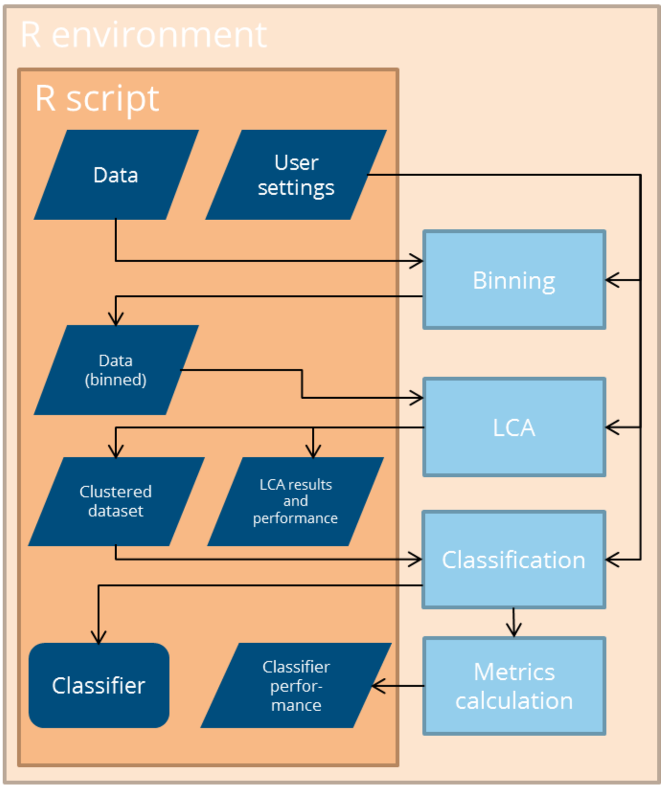


Figure 3.12: Overview of the implementation in R

The R programming environment, version 3.2.3 released in August 2015, is the foundation of the implemented algorithms. The implementation consists of four main functions, shown in the light blue blocks on the right of figure 3.12. These functions are all defined in their own .R file. A script forms the basis of the implementation that calls the various required functions. In addition, the script handles loading the data that is to be used and defines the settings for function calls. A short overview of the script is given here, but a more detailed overview of the script and the packages that have been used can be found in Appendix C.

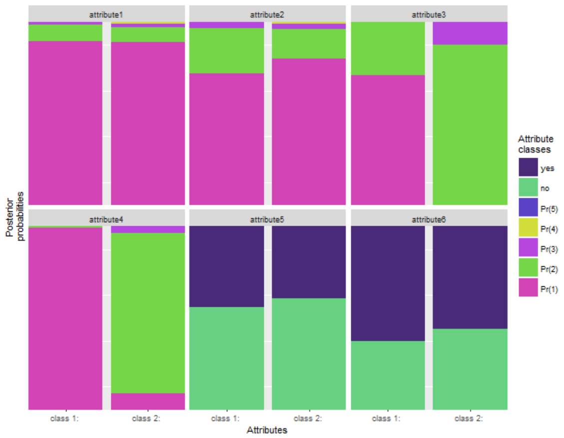
The first step is loading the data that contains the passenger and behavioural attributes that are to be used. The data is loaded from a CSV-file, although other formats are also possible as long as they result in a dataframe in the R environment. Because the LCA package, poLCA, does not support continuous variables, these have to be binned. This is done using the binning function.

After the continuous variables have been binned, the data set contains only categorical variables. Consequently, the data can be used in the LCA function. This LCA function requires a formula that defines which of the attributes in the data set should be used to form the classes. In the present case, these are the passenger attributes. The LCA function outputs the data set along with the classes. Additionally, the function returns the information about the fit of the model in the form of the BIC and AIC. Detailed information about the classes that were found is also given, such as the class-conditional item response probabilities and the estimated class population shares. Examples of these outputs are shown in figure 3.13a and c.

The data set with the cluster labels can now be used to create the classifier. For this, the classification function is used. To use this function, the attributes that should be used for the classification are defined. These are the passenger attributes. After the classification has been done, the metrics calculation function is called in order to calculate the performance metrics for the classifier. The performance, as well as some properties of the classifier is printed on the screen. Examples of the classification output can be seen in figure 3.13b and d. The actual classifier remains in the R environment and can be interacted with using the functions in the adabag package.

```
=========================================================
Fit for 2 latent classes:
=========================================================
number of observations: 2097
number of estimated parameters: 27
residual degrees of freedom: 692
maximum log-likelihood: -7835.366

AIC(2): 15724.73
BIC(2): 15877.23
G^2(2): 486.1154 (Likelihood ratio/deviance statistic)
X^2(2): 867.4241 (Chi-square goodness of fit)
```

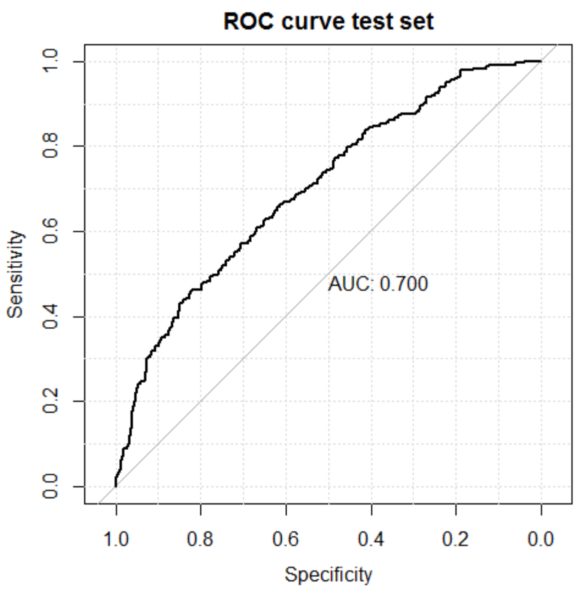
(a)



(c)

```
==========================================================
Classification Results:
==========================================================
Confusion matrix for training set:
              Predicted Classes
Actual classes Class  1 Class  2
      Class  1      722       93
      Class  2      134      449
Overall error on training set: 0.1623748
Average error on training set: 0.1623748
Macro F1-score on training set: 0.83196

Confusion matrix for testing set:
              Predicted Classes
Actual classes Class  1 Class  2
      Class  1      307       96
      Class  2      141      155
Overall error on test set: 0.3390558
Average error on test set: 0.3390558
Macro F1-score on test set: 0.6470291

Relative importance of predictor variables
    variable importance
1 predictor1   45.91308
2 predictor3   18.89595
3 predictor2   17.59633
4 predictor4   17.59464

***ROC Curve Results (training)***
Area under the curve: 0.9346

***ROC Curve Results (test)***
Area under the curve: 0.6999
```

(b)



(d)

Figure 3.13: Output examples for LCA (a and c) and classification (b and d) in R

3.6 Chapter conclusion

Based on the findings of chapter 2, this chapter has presented requirements to the input, output, and performance of behavioural classification. Based on these requirements, the CC-framework has been presented. This framework consists of two main parts. First, based on passenger attributes, it performs clustering. This forms behavioural classes on a training data set. Second, based on the clustered data set and the passenger attributes of the training data set, it creates a classifier. After this classifier has been created, 'new' passengers can be classified into one of the behavioural classes by the classifier. The behavioural characteristics associated with the behavioural class form the input for the behavioural model in the PDF.

For both parts of the framework, requirements have been set and techniques have been introduced. Based on these requirements, a specific technique for both parts of the framework has been chosen and implemented. For clustering, Latent Class Analysis has been chosen because it offers several benefits; LCA is able to handle mixed-type data, has been shown to yield good results, and provides good performance metrics due to being model-based. LCA has been implemented in the R programming environment using the poLCA package. For the classification part of the CC-framework, a boosted ensemble classifier based on decision trees has been chosen. This solution was chosen because it offers a transparent classifier, supports mixed-type data, and decision trees are relatively quick to compute. However, because decision trees are less accurate than some other algorithms, a boosted ensemble classifier has been used in order to increase accuracy. The boosted ensemble classifier has also been implemented in the R using the adabag package. Both the clustering and classification implementation have then been integrated into one in R.

In the next chapter, the implementation of the CC-framework will be applied to a data set in order to test its performance. Because there are various parameters in the framework that can be adjusted, a grid search for these parameters will be performed to find the optimal settings.

4

Clustering and Classification Results

The previous chapter has introduced various methods for clustering and classifying data into groups and finally presented an integrated implementation of the chosen methods. The present chapter will go more into depth about the application of the framework. To this end, the framework is applied to a data set in order to test its performance. Based on this, some conclusions with respect to the behavioural classes of these data and the performance of the framework can be made.

This chapter is structured as follows. First, a description of possible sources of data and the actual data set that is used with the implemented CC-framework is given in section 4.1. The data set is presented using descriptive statistics in section 4.2 in order to discover possible relations between variables in the data. Based on this, our expectations with respect to the results of classification and clustering can be established. Subsequently, the optimal values for some parameters for clustering and classification are sought for in section 4.3. Detailed results of clustering and classification on the data set based on these optimal parameter values are presented in section 4.4. The findings in this chapter are summarised in 4.5.

4.1 Description of used data

AAS has several systems in place that could potentially provide data that can be used to test the CC-framework with. However, in all cases, passenger privacy is a main concern. For this reason, available data is anonymised and available passenger characteristics are stripped from the data. As a result, there are no personal characteristics available in these data. As established earlier in this report, because both passenger characteristics and behavioural characteristics are required for classification, these data are hence unsuitable to use in this thesis. Still, there are two interesting data sets that contain data acquired from RF-positioning. These two sets are discussed in sections 4.1.1 and 4.1.2. One other data set contains both passenger and behavioural characteristics, and multiple categories of these characteristics, i.e.: personal, process, and trip characteristics. This data set, which is the PASSME data set, has therefore been used with the framework. However, contrary to the other two aforementioned data sets, the PASSME data set contains no sensor information, but only survey data. The PASSME data set is introduced in section 4.1.3.

4.1.1 Bluetooth and Wi-Fi detection

AAS has deployed a rather extensive network of combined Bluetooth and Wi-Fi tracking sensors, called the BlipTrack system.[4] Each mobile device that has Wi-Fi or Bluetooth enabled that comes into the range of such a sensor is registered. Combining the observations of the various sensors that have registered a device can in theory lead to quite a detailed overview of the behaviour of a passenger

[4] BlipTrack is a product of BLIP systems, based in Denmark.

with respect to the amount of time spent in areas of the airport and the route taken. As the implementation of BlipTrack at AAS is primarily aimed at lead time prediction, the system is quite dense at passenger process areas such as check-in desks and security filters. For example, the check-in desk area of departures 1 is equipped with 17 of such sensors. Also the security filter 1 and lounge 1, which follow after departures 1, are equipped with the sensors. In theory, using these data could provide quite a complete picture of the locations a passenger has visited, and for how long.

However, the data have a few limitations to use them for behavioural classification. First, although the BlipTrack system can be used as a tracking system, the implementation at AAS is primarily aimed at lead time prediction and tracking area occupation. This results in a lower level of detail in the data. Combining the raw data still leads to information about entering and leaving the check-in area, security area, and lounge area. Second, the data are anonymised and there are no passenger attributes available. This means that for every tracked device in the data set, the only data available are the timestamps at which it was registered by the sensors and the hashed MAC-address of the device. Consequently, it is not possible to couple the BlipTrack data to any other systems, should these have more information coupled to a specific device's MAC-address. For these reasons, the BlipTrack data was considered unsuitable for usage in the CC-framework.

4.1.2 Self-Service Boarding Pass Check

Passengers that want to proceed from the departure hall to the security filters at AAS are required to scan their boarding pass at the self-service boarding pass check (SSBPC). The SSBPC checks if the boarding pass is valid and, if this is the case, grants the passenger access to the security filter area. The SSBPC-system records various aspects of each scanned boarding pass, such as the access time, travel class and flight number. Theoretically, this system could also provide information about the passenger, such as name and sex. However, due to privacy regulations, these data are stripped from the system. What remains is a rather limited set of information for each individual passenger that passed the SSBPC that almost exclusively pertains to the flight of the passenger and not the passenger itself. While some attributes do in fact affect behaviour according to the findings in section 2.2.2, the only attribute in this data that can be characterised as a behavioural attribute is the time remaining between scanning the boarding pass and the departure of the flight, which is not influenced by the other attributes in the data according to the same findings.

However, based on some of the attributes in the SSBPC data, some additional attributes can be added to the data. Using the IATA code of the destination airport, it can be determined whether or not the destination is within the Schengen Area. Additionally, based on the coordinates of the destination airport, the great circle distance of the flight can be calculated. This can be used as a proxy for flight duration, which may affect how long before their flight passengers enter the security filter. Additionally, the exact time in minutes to departure can be calculated based on the date and time of passage and the date and time of flight departure. However, even with these added attributes, the data set is still very limited with respect to both behavioural attributes and passenger attributes. Consequently, the usability of the data for the envisioned goal of finding and classifying behavioural groups is limited.

4.1.3 PASSME survey data

The PASSME data set consists of surveys taken from departing and transferring passengers at an airport within the PASSME context. The data set contains 3,923 respondents, out of which 2,097 are departing passengers and 1,826 are transferring passengers. Although there are no specific requirements with respect to the sample size for clustering and classification, this sample size is quite large and expected to suffice. The two subsets predominantly have the same attributes. Several of these have been shown to be related, according to the literature review in section 2.2.2. Because there are several attributes unique to either departing or transferring passengers, and the passenger process for these two types of passengers differs, the two subsets will be treated as two separate data sets.

Table 4.1 describes the attributes in the total data set. The column 'category' corresponds to the three categories of characteristics described in section 2.2.1.2. The column 'type' indicates if the attribute is either of type 'passenger' or 'behavioural', corresponding to the definition as introduced in section 2.2.1.1. Recall from chapter 3 that clustering is performed based on the behavioural attributes, yielding the behavioural classes. Using the passenger characteristics of all observations in these classes, classification rules are made.

The last column of table 4.1 indicates the variable type, which could be either categorical or continuous. Recall from section 3.3.5 that continuous variables that are to be used in the LCA, i.e. behavioural attributes, should be binned first in order to convert them to a categorical type.

Table 4.1: Attributes in the PASSME data set

Common attributes				
Attribute	**Value**	**Category**	**Type**	**Variable type**
Age group	00-30, 31-60, 60+	Personal	Passenger	Categorical
Country of residence	Within or outside European Union	Personal	Passenger	Categorical
Destination country	Within or outside European Union	Trip	Passenger	Categorical
Flight day of week	Monday through Sunday	Trip	Passenger	Categorical
Flight frequency	0-3, 4-10, 11+, per year	Personal	Passenger	Categorical
Nationality	Within or outside European Union	Personal	Passenger	Categorical
Passenger amount	Amount of passengers	Process	Passenger	Categorical
Travel class	Economy, business/first, unknown	Trip	Passenger	Categorical
Travel duration	1, 2, 3, 4-7, 8-13, 14-21, 21+, days	Trip	Passenger	Categorical
Travel purpose	Business or leisure	Trip	Passenger	Categorical
Went shopping	Indicates whether the person has *bought* anything in a shop	Process	Behavioural	Categorical
Went to restaurant	Indicates whether the person has *bought and consumed* anything in a restaurant/coffee stand/etc.	Process	Behavioural	Categorical
Departing passenger attributes				
Attribute	**Value**	**Characteristic type**	**Attribute type**	**Variable type**
Check in hall	1, 2 or 3	Process	Passenger	Categorical
Transport mode	Travel mode to the airport	Trip	Passenger	Categorical
Travel time to the airport	Travel time from home to the airport, minutes	Process	Passenger	Continuous
Landside time	Time spent landside, minutes (area between the airport entrance and security check)	Process	Behavioural	Continuous
Lounge time	Time spent in lounge, minutes (area after security check, but before the piers)	Process	Behavioural	Continuous
Go to gate time	Time between going to gate (from lounge) and scheduled flight departure, minutes	Process	Behavioural	Continuous
Gate time	Time spent at gate, minutes	Process	Behavioural	Continuous
Total time at the	Total time between arriving at	Process	Passenger	Continuous

airport	the airport and flight departure			
Transferring passenger attributes				
Attribute	**Value**	**Characteristic type**	**Attribute type**	**Variable type**
Total transfer time	Total transfer time, minutes	Process	Passenger	Continuous
Lounge time	Time spent in lounge, minutes	Process	Behavioural	Continuous
Go to gate time	Time between going to gate (from lounge) and scheduled flight departure, minutes	Process	Behavioural	Continuous
Gate time	Time spent at gate, minutes	Process	Behavioural	Continuous
Origin country	Country where transferring passenger departed from	Trip	Passenger	Categorical

The attribute type of most attributes in the set is quite unambiguous. Yet, the attribute type of some attributes in table 4.1 might intuitively seem strange. Some of them are therefore explained here:

- **Flight frequency** may be interpreted as behaviour and to an extent it certainly is, however it is on a different scale. In the present case we are interested in the behaviour of passengers within the terminal, not in their behaviour on a yearly basis. Flight frequency on a yearly basis is hence not a behavioural attribute, but rather a passenger attribute that may explain the behaviour of a passenger within the airport. After all, an experienced passenger who flies several times per year may display different behaviour than a passenger who flies once per year or less.
- **Travel time to the airport** is a component of behaviour on the strategic level. However, with respect to behaviour within the terminal, this attribute is exogenous and hence regarded as a passenger attribute.
- **Landside time** indicates the amount of time the passenger spends at landside, i.e. the amount of time between entering the terminal and joining the security filter queue. A passenger may choose to perform discretionary activities during this period, influencing his landside time. Hence, landside time is represented as a behavioural attribute.
- **Total time at the airport** is also a consequence of the departure time choice and mode choice of the passenger, which is exogenous to the behaviour at the airport. However, the total time at the airport can affect the activities that a passenger will perform at the airport and is hence interpreted as a passenger attribute.
- **Total transfer time** is exogenous to the behaviour of the transfer passenger at the airport as it has been established when the flight was booked. However, the total time available at the airport may affect the behaviour of the passenger at the airport and is hence regarded as a passenger attribute.

Because the PASSME data is based on surveys, and not collected specifically for the purpose of this thesis report, the data is somewhat lacking with respect to the amount of detail and the number of attributes included. For example, the definitions of all behavioural attributes are quite broad. Comparing the PASSME data with the characteristics that have been shown to be related to behaviour in section 2.2.2, some of these attributes are present in the PASSME data. Below is an overview of these attributes, and their effects on behaviour according to section 2.2.2:

- **Age** has been shown to be related to the likelihood of shopping.
- **Total time at the airport** has been shown to be related to the likelihood of food and drinks consumption.
- **Travel class** has been shown to be related to the waiting time in the three departing time phases. These phases are not specifically present in the PASSME data, although they do contain the time spent in different parts of the airport terminal.

- **Travel destination** has been shown to be related to the likelihood of consuming food or drinks. However the PASSME data discerns between travellers with a destination within or outside the EU, instead of intercontinental and continental travel. However, these definitions are fairly similar.
- **Travel experience** has been shown to be related to the likelihood of shopping and performing discretionary activities before security. However, the latter is not specifically available in the PASSME data. Moreover, travel experience is not available in the PASSME data, though this could be approximated by the 'flight frequency' attribute.
- **Travel purpose,** for which it has been shown that business passengers are less likely to perform discretionary activities for the security check. Business passengers also allow for a larger safety margin with respect to their arrival at the airport. Both of these attributes are not specifically available in the PASSME data, though they could be approximated by the 'shopping', 'restaurant', 'landside time', and 'total time at the airport' attributes.

4.2 PASSME data analysis

Using a latent class analysis on the PASSME data set will form a specified number of classes for which the model fit can be verified using metrics such as the AIC and BIC. It is however important to realise what kind of classes can be expected in order to be able to also logically verify the validity of the classes that are formed. In addition, it is useful to explore possible predictor variables for the classification. Therefore, the data in the data set will be discussed and analyses will be performed in order to find possible relations between attributes in the data.

In the previous section, the attributes in the data set have been categorised into behavioural attributes and passenger attributes. The behavioural attributes will be used to form the classes, while the passenger attributes will be used to classify passengers into these classes. We are therefore mainly interested in the following two questions:

1. How are the behavioural attributes distributed, and can groups be observed based on these distributions?
2. Are there relations between passenger attributes and behavioural attributes?

Subsection 4.2.1 pertains to the first question. Here, the six behavioural attributes are analysed by exploring the distributions of the numerical variables and possible correlations between them using Pearson correlation. In addition, it is tested if the distributions of the numerical behavioural attributes differ significantly across the categories of the two categorical behavioural attributes in the set using the Mann-Whitney U test.

The second question is discussed in subsection 4.2.2. First, the distributions of the travel purpose and age over the days of the week are analysed. Significant differences over the days of the week for these attributes may be related to differences in behaviour, as these attributes have been shown to be related to behaviour (see section 2.2.2). Next, the transport mode and flight destination are shown in relation to the total time spent at the airport. Lastly, possible relations between the passenger attributes and behavioural attributes are tested.

Note: due to the confidentiality of the PASSME data, some text and figures regarding the data analysis and results of the framework had to be omitted for this public version of the report. Omissions are indicated in the text.

4.2.1 Analysis of behavioural attributes

This subsection has been omitted.

4.2.2 Observations with respect to passenger- and behavioural attributes

This subsection has been omitted.

4.2.3 Summary

This section has presented various statistics and plots of the PASSME data set. This overview is by no means all-encompassing. However, to the best of the author's knowledge, the most interesting combinations of attributes were discussed. Summarizing, this section has shown the following:

- Some of the numerical behavioural attributes are (slightly) correlated. Time in lounge is correlated with all three other numerical attributes, and the go to gate time and gate time are correlated. Additionally, the distribution of the numerical attributes in many cases differs significantly across the categories of the categorical behavioural attributes.
- There is some variance in the percentage of business travellers, with a peak after the middle of the week. The percentage of business travellers is higher for transferring passengers than for departing passengers.
- As the total time spent at the airport increases, the percentage of travellers that choose to either shop or visit a restaurant increases.
- *Omitted.*
- There is not much difference in the total time at the airport for the various ingress transport modes; only the mode 'coach travel organisation-charter bus' shows a significant difference compared to some of the other modes. However, passengers with a destination outside the European Union spend more time at the airport compared to passengers with a destination within the European Union.
- As the total time at the airport increases, the proportion of time spent in the lounge increases. For departing passengers, the proportion of time spent landside also increases. For both departing and transferring passengers, the percentage of time spent at the gate is the highest until about three hours dwell time at the airport. After that, more time is spent in the lounge, or landside.

Based on the analyses of the behavioural attributes, it can be mentioned that the differences in the shapes of the distributions, the correlation between the various numerical attributes, and the significant differences for the distributions across the classes of the categorical variables, lead to expect that indeed classes could be found in the data. It is likely that there will be a visible difference between shopping and restaurant behaviour across the classes that are to be found. In addition, several possible class predictor variables have been identified, such as travel purpose or travel experience as predictors for the time spent in the airport lounge.

4.3 Optimising the Clustering and Classification parameters

When executing the implemented CC-framework, there are quite some parameters that affect the performance of both the clustering and the classification. Hence, there are essentially two models that should be optimised. In principle, clustering is the first part of the process. It makes sense to first find the best clustering results, and then apply classification based on this clustering. However, the best clustering results do not necessarily lead to the best classification result. This could lead to a well-fitting division of clusters, but with clusters for which only a mediocre classification can be made that is not better than random guessing, such as in the case of AUC < 0.5. Because the goal of this thesis work is most of all to be able to classify passengers, it was chosen to adopt an integrated approach wherein both clustering and classification are performed and the results of both are assessed. Using a grid search, the parameters of clustering and classification are varied in order to find the best parameters settings for optimal classification performance.

During the grid search, three key inputs of the CC-framework are varied:

- The **number of classes** to be used in the latent class cluster analysis.
- The **bin size** of bins that are used for the time attributes in the data set.[5]

[5] Recall that the LCA can only deal with categorical attributes, hence binning the time attributes of the PASSME data is required

- The **maximum tree depth** of the decision trees that are used in the SAMME ensemble

Table 4.2: Parameter ranges for classification optimisation

Parameter	Minimum	Maximum	Step size
Number of classes	2	7	1
Bin size	5 minutes	60 minutes	5 minutes
Maximum tree depth	10	15	1

For both subsets of the PASSME data set, these three parameters were varied as shown in table 4.2, which were chosen based on results of preliminary tests to find sensible parameter ranges. All possible combinations within these ranges have been tested, leading to a total of 432 results. Each result contains the BIC and AIC for the clustering part. The metrics for classification are recorded separately for the training set and the test set. Classification metrics that have been used are the AUC, overall accuracy, average accuracy and the F1-score[6]. Recall that the performance of the classifier on the training set is expected to be better than the performance on the test set. Therefore the metrics for the test set are leading with regards to choosing the best parameter settings.

In the following two subsections, the results for the departing and transferring subsets of the PASSME data are presented and explored. Based on the results, a choice with respect to the parameter settings will be made. The next section, section 4.4, will present more in-depth results of the chosen parameter settings.

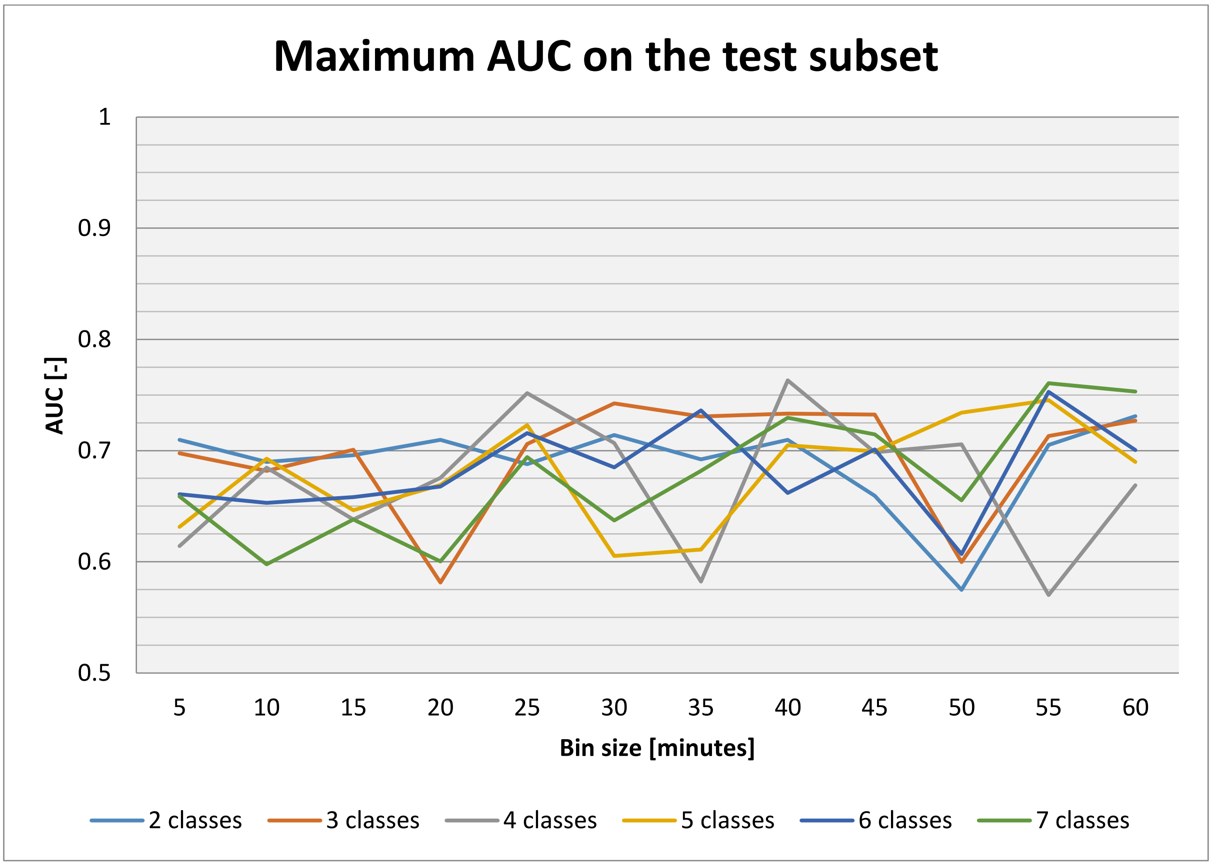
4.3.1 Departing passengers

Figure 4.1 shows the results for the subset of departing passengers. The graphs show the performance per class, with the bin size for the numerical behavioural attributes on the horizontal axis. Because this leaves out the dimension of the maximum tree depth, the maximum performance with respect to this parameter is shown. This is acceptable because there is little difference in results between the different settings for maximum tree depth. Larger versions of the graphs and some additional graphs are included in Appendix E.

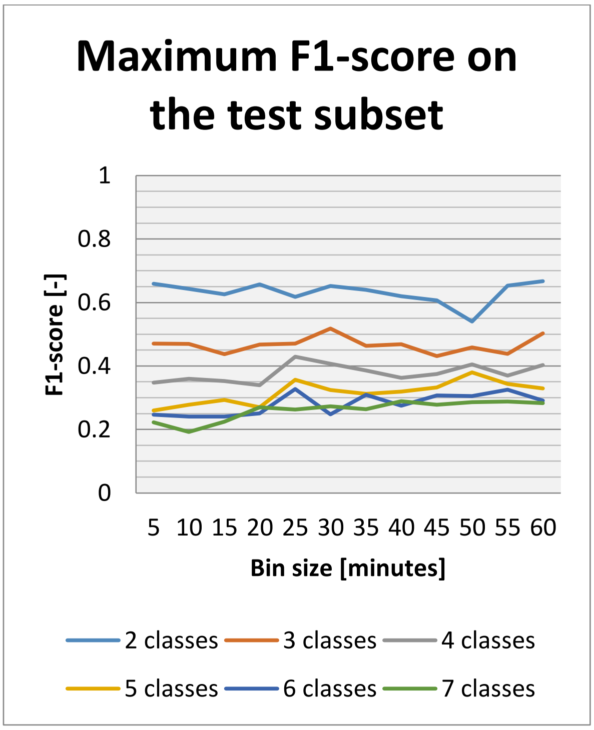
The AUC values in figure 4.1a appear to be rather variable. However, a few observations can be made. The absolute maximum AUC of 0.76 is found at four classes and a bin size of 40 minutes. The AUC-values for all classes tend to move closer together at bin sizes of 25 and 40 minutes.

The average accuracy, included in the appendix, paints a different picture. The accuracy increases as the number of classes increases. Specifically for the two and three class case, the accuracy increases as bin size increases. Examining the overall accuracy, the picture is almost reversed; here, the accuracy decreases as the number of classes increases. The difference between the overall accuracy and average accuracy can be explained by the fact that the average accuracy does not compensate for the class size. The F1-scores show an increase in score as the number of classes decreases. There is no apparent effect of bin size on the F1-score.

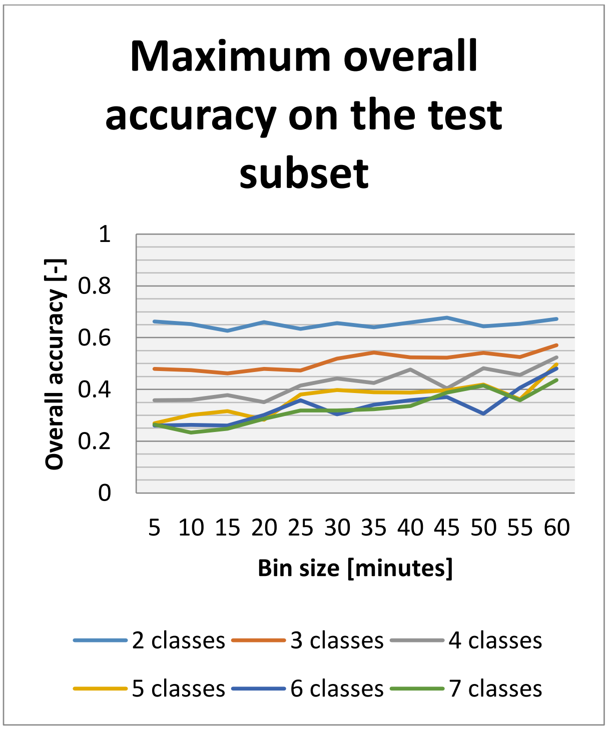
[6] The F1-score is the F-score with a β value equal to one



(a)



(b)



(c)

Figure 4.1: Optimisation results for the departing passengers subset

In addition to the metrics of the classification, the model fit of the latent class clustering should also be assessed. The AIC and BIC are shown in figure 4.2, summarized as the minimum result of all bin sizes that were tested for each amount of classes. The BIC value increases as the number of classes increases, though there is a slight dip at four classes. Consequently, according to the BIC, a clustering with four classes yields the best model. The AIC contradicts this, with the value decreasing as the number of classes increases. This may be attributed to the fact that the AIC tends to favour overly complex models (Dziak et al., 2012). With respect to bin size, it can be noted that both AIC and BIC decrease for larger bin sizes.

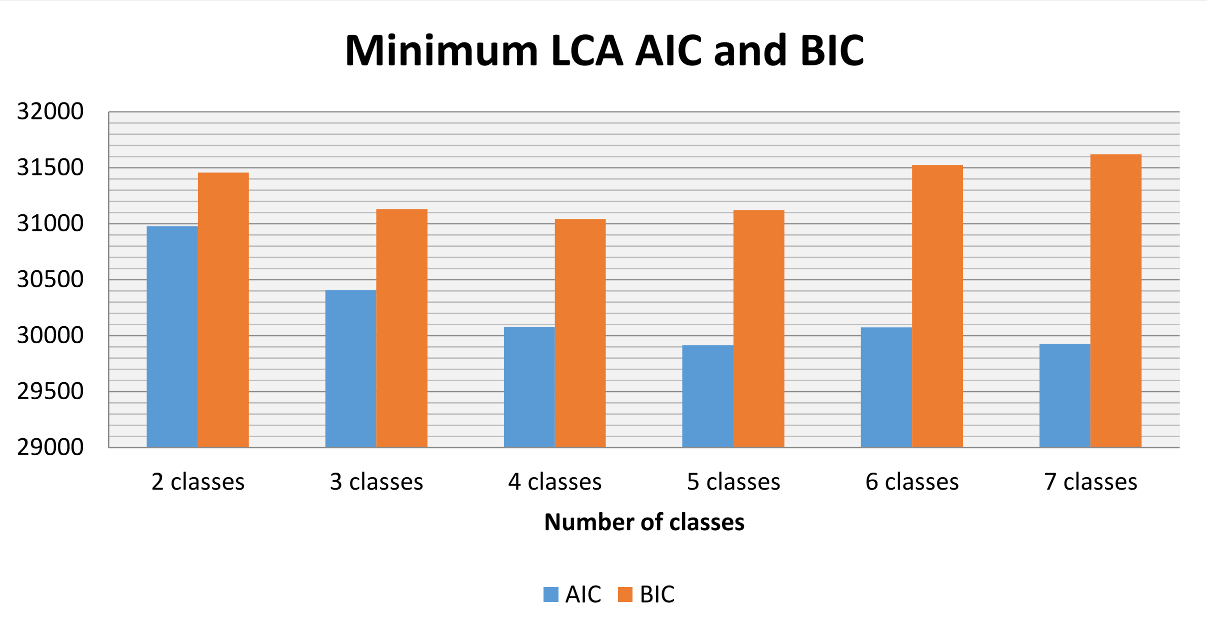


Figure 4.2: Cluster model performance for various bin sizes for the departing passengers subset

Summarizing the presented results for clustering and classification, the following remarks can be made:

- Two of the classification metrics favour a low number of classes, one favours a high number of classes, and one is not clear.
- The LCA model fit is best for around two to four classes.
- The maximum tree depth parameter has little effect on any of the results.

Based on these observations, the best fit for both clustering and classification is achieved with two classes. For two classes, the best performance on all classification metrics is achieved with a bin size of 60 minutes. This has led to the values presented in table 4.3.

Table 4.3: Optimal parameter values for the departing passengers subset

Parameter	**Optimum**
Number of classes	2
Bin size	60 minutes
Maximum tree depth	14

4.3.2 Transferring passengers

The results of optimisation on the subset of transferring passengers are quite similar to the results of the departing passengers. Some figures are included here. For a complete overview of the results, the reader is referred to Appendix E.

Generally, performance of the classifier on the test set increases as the bin size increases and the number of classes decreases. Looking at the AUC values in figure 4.3, the peaks for two classes at 45 and 55-minute bin sizes stand out, as well as the result for seven classes at a bin size of 60 minutes. However, taking into account the overall accuracy, average accuracy and F1-score, the same picture is painted as for the departing passenger subset. Hence, there is a higher F1-score and overall accuracy for a low number of classes, whereas the average accuracy is higher for a high number of classes.

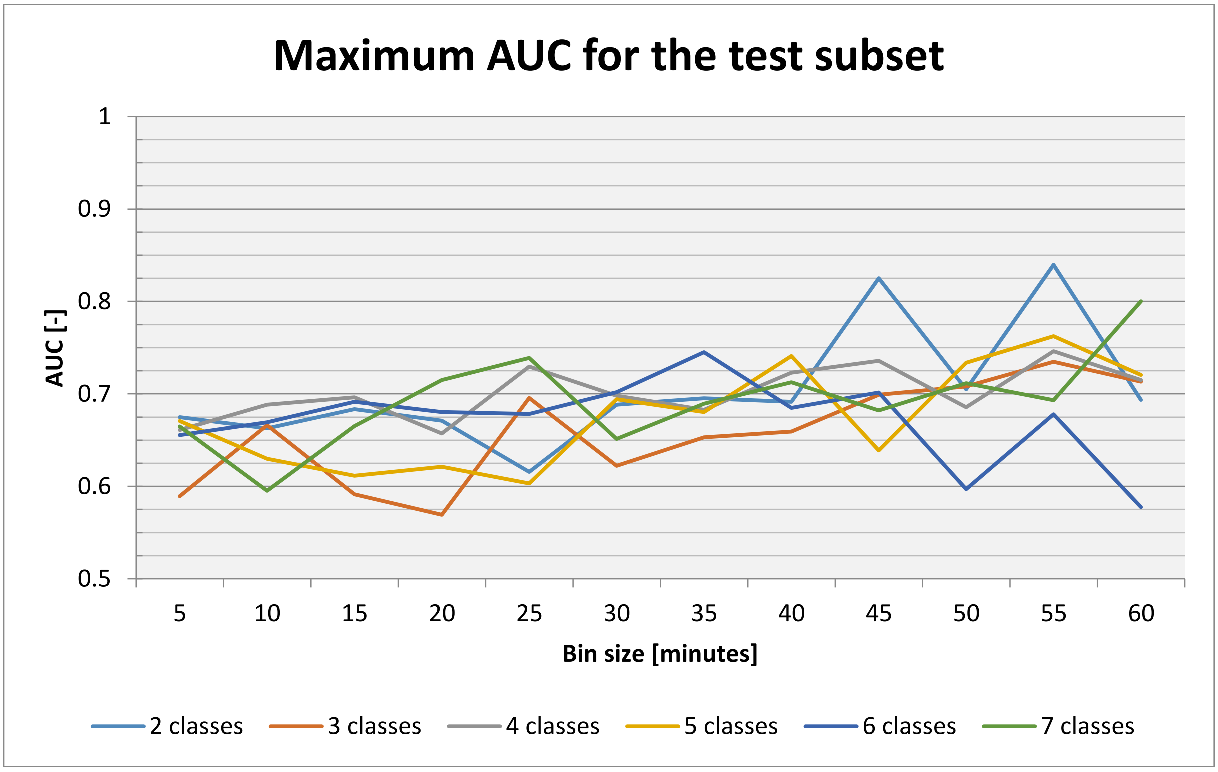


Figure 4.3: AUC results for the transferring passengers subset

The LCA model BIC and AIC decrease as the bin size increases, similar to the departing passenger subset. However, with respect to the number of classes, the BIC and AIC have their optimal value at four and seven classes, respectively. Once again, this may be explained by the fact that the AIC tends to favour more complex models.

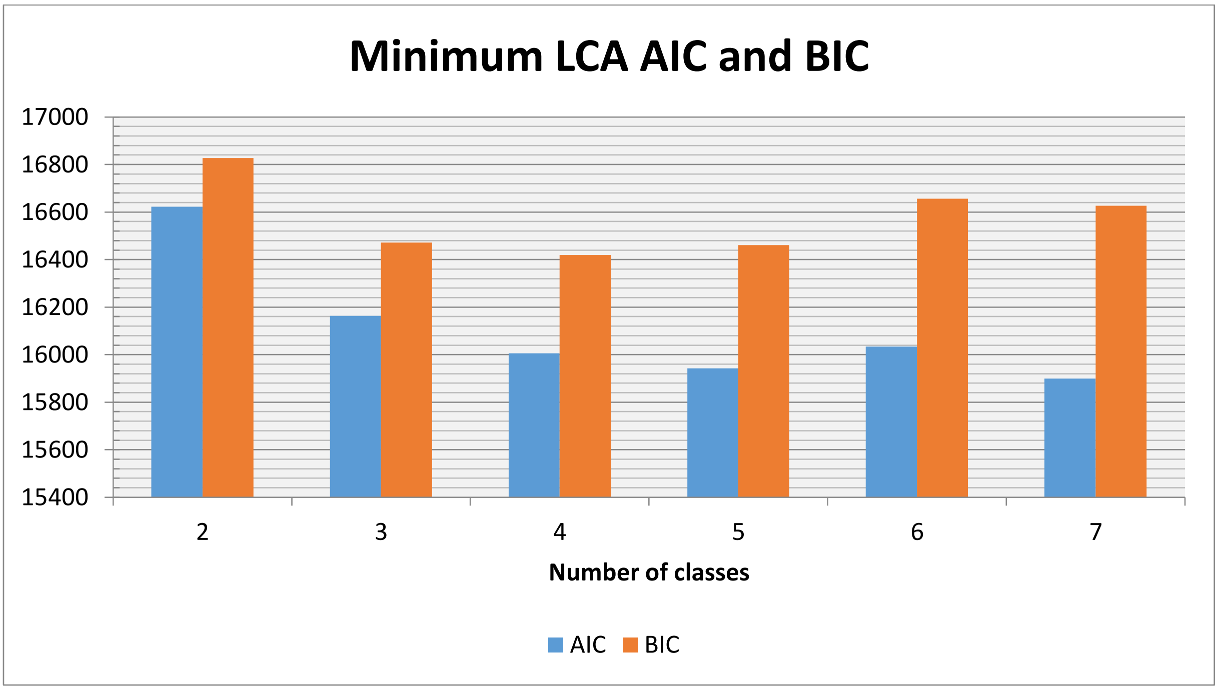


Figure 4.4: Cluster model performance for various bin sizes for the transferring passengers subset

Based on the results for the transferring passengers subset, some remarks can be made:

- With respect to classification, a low number of classes and a large bin size lead to the best classifier performance.
- The LCA model yields the best results for around four to five classes. However, classifier performance for this number of classis is much worse.
- Again, the maximum tree depth parameter has little effect on any of the results.

Based on these observations, the parameter values for transferring passengers will be set as shown in table 4.4.

Table 4.4: Optimal parameter values for the transferring passengers subset

Parameter	Optimum
Number of classes	2
Bin size	45 minutes
Maximum tree depth	10

4.4 Final results

The number of classes, bin size, and maximum tree depth settings for the CC-framework have been optimised for the PASSME data set in section 4.3. In this section, results are presented in further detail. There are some random components in the framework, for example due to the sampling of the training subset for the ensemble classifier. This can lead to slightly different results between different runs of the script. Because of this, the CC script has been run 50 times in order to give confidence intervals for the performance metrics. Results with respect to the class distributions, class probabilities and predictor importance are difficult to verify over different runs as the class labels do not remain same over different runs (e.g.: class 1 during the first run could be class 2 during the second run). As such, these results are displayed for one run of the script, though with a fixed seed in order to guarantee reproducibility.

4.4.1 Departing passengers

The results for the subset of departing passengers were obtained using the six behavioural attributes as defined in table 4.1 for the LCA. For the ensemble classifier, the thirteen passenger attributes as defined in the same table were used. An overview of this is given in table 4.5.

Table 4.5: Attributes used in the LCA and classification for departing passengers

LCA	Classification	
Landside time	Age	Flight frequency
Lounge time	Booking class	Nationality
Go to gate time	Check-in hall	Passenger amount
Gate time	Country of residence	Total time at the airport
Shopping	Day of week	Transport mode
Restaurant	Destination Country	Travel duration
	Travel purpose	

The total number of objects in the subset of departing passengers is 2,097. The fixed result of the LCA has 1,218 (58%) in class 1, and 879 (42%) in class 2.

4.4.1.1 Classes

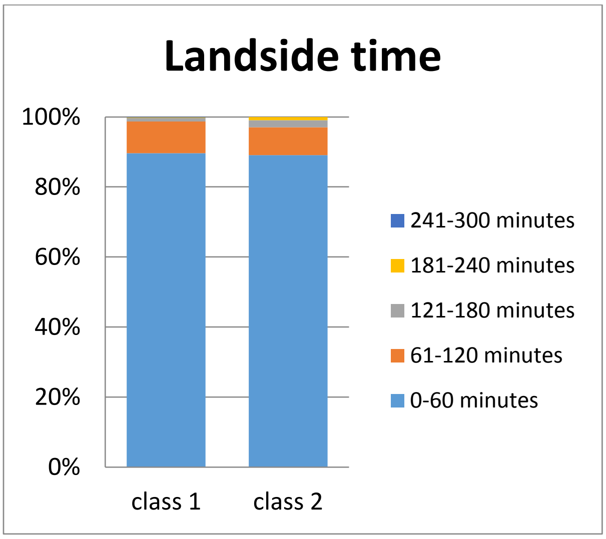
The LCA has been performed 50 times. The model fit, expressed by the BIC and AIC values, shows very little variation as shown in table 4.6. This means that there is little difference in the goodness of fit between the various model estimations.

Table 4.6: LCA performance for departing passengers

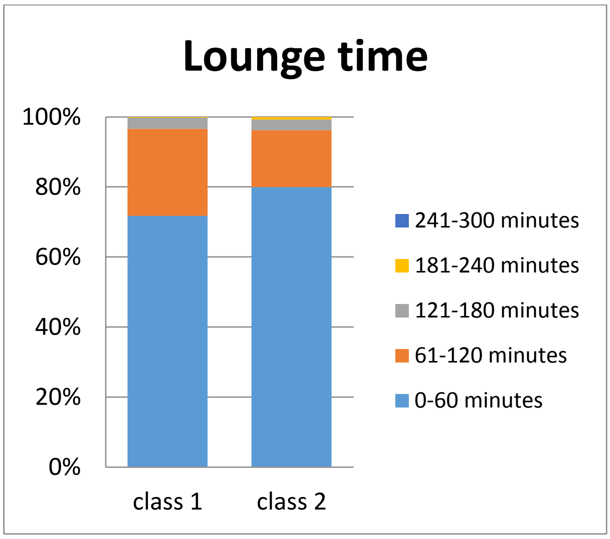
Metric	Mean	95% confidence interval
BIC	15877.45	± 0.17
AIC	15724.95	± 0.17

Figure 4.5 shows the estimated class-conditional outcome probabilities of the latent class analysis. For each attribute used in the LCA, a bar graph is shown for each of the two classes. Each colour on the bar represents the probability that an object classified in that class will have the outcome represented by that colour. From the figure it becomes apparent that the go to gate time and the time spent at the gate are the main difference between the two classes. The vast majority of class 1 (70%) goes to their gate 60 minutes or less in advance of flight departure. For class 2, this percentage is almost zero; by contrast, almost 88% percent of class 2 goes to their gate 61 to 120 minutes before their flight departs. The same can be seen for the time spent at the gate. For class 1, 99% spends less than an hour at the gate, while class 2 has an 87% probability of spending 61 to 120 minutes at the gate. The other attributes are very similar between the two classes, though class 1 has a slightly higher probability of spending more time in the lounge.

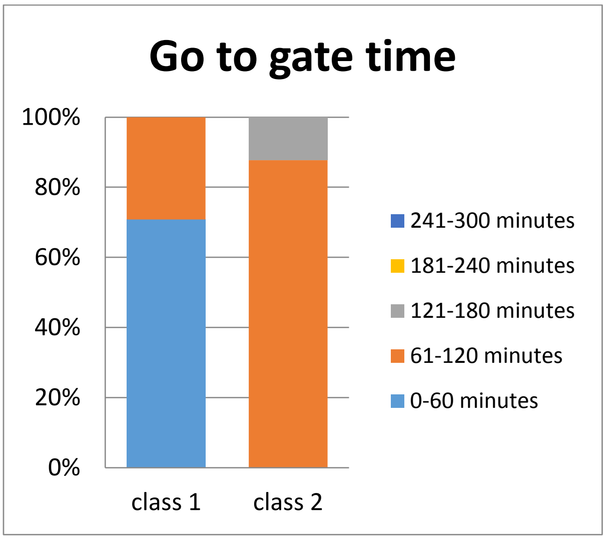
Consequently, if one has to characterise the two classes, class 1 could be described as the 'late-at-the-gate' class. Class 2 could then be described as the 'early-at-the-gate' class.



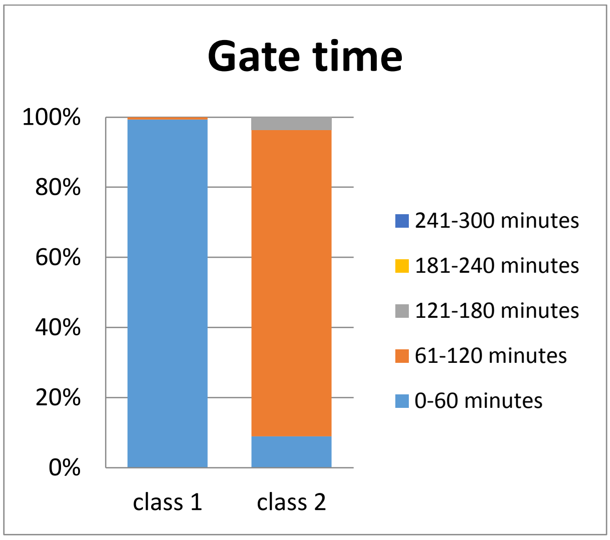
(a)



(b)



(c)



(d)

Figure omitted

(e)

Figure omitted

(f)

Figure 4.5: LCA result for departing passengers

The presented classes are primarily distinguished based on behaviour with respect to going to and spending time at the gate. It is interesting to note what happens if these attributes are left out of the LCA. For example, leaving out the time spent at gate in the LCA leads to two classes that are primarily distinguished with respect to the time spent in the lounge. At the same time, the class that spends the most time in the lounge also has a considerably higher probability for shopping and restaurant visit. In addition, the goodness of fit of the LCA and the performance of the classifier are higher compared to when the time spent at the gate is included in the model, implying that this results in a better model.

4.4.1.2 Classifier

Table 4.7 shows the mean and 95% confidence interval of the performance metrics for ensemble classifier, based on 50 model estimations. In all cases, the metrics for the training set are better than for the testing set. This is not unexpected as the classifier is created using the training set, hence creating better results for this set. Note that the overall and average accuracy are equal because there are only two classes.

Table 4.7: Classifier performance for departing passengers

Set	Metric	Mean	95% confidence interval	
train	AUC	0.9787	±	0.0016
test	AUC	0.6883	±	0.0049
train	Overall accuracy	0.9163	±	0.0040
test	Overall accuracy	0.6363	±	0.0037
train	Average accuracy	0.9163	±	0.0040
test	Average accuracy	0.6363	±	0.0037
train	Macro F1-score	0.9141	±	0.0041
test	Macro F1-score	0.6250	±	0.0039

Overall, the value of all metrics is around 0.65, indicating a relatively poor performance of the classifier. Judging from the values in table 4.8, the classifier mainly struggles to recognize objects that should be categorized in class 2.

Table 4.8: Per-class recall and precision for departing passengers

Class	Recall	Precision
Class 1	0.730	0.698
Class 2	0.571	0.608

With respect to the importance of the predictors used in the classifier, shown in figure 4.6, it can be noted that the five most important predictors out of the total of thirteen make up for 77% of the total prediction of the classifier. The total time spent at the airport dominates as a predictor and is almost twice as important as the next most important predictor, which is the day of the week. It is however logical that the total time at the airport has such a high importance, seeing as four of the six attributes used in the classification pertain to time, of which this attribute is the total. However, leaving the total time at the airport out of the ensemble classifier drops the performance of the classifier by only a few percentage points. Hence, there is some correlation with other predictors.

Other relatively important predictors, each having an importance higher than ten percent, are the day of the week, the transport mode that was used for transport towards the airport, and the total duration of the journey.[7] This is in line with the observations of 4.2.2, which has shown significant differences between the duration of staying at the airport for the different transport modes and the difference between passenger types over the days of the week.

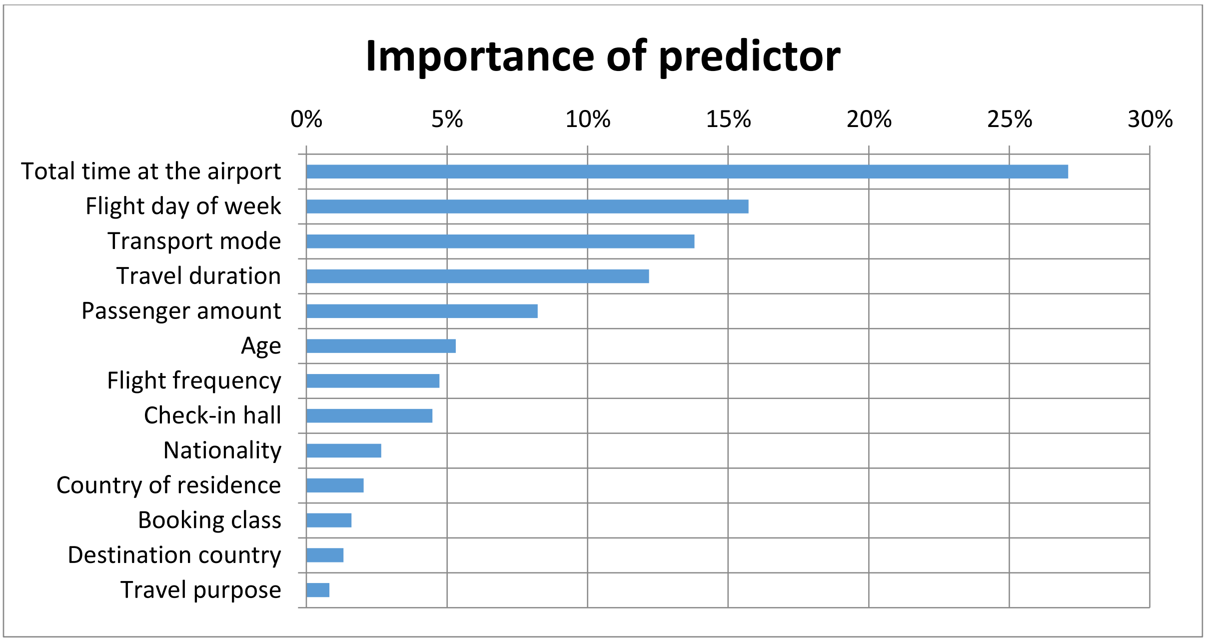


Figure 4.6: Importance of predictor attributes for departing passengers

4.4.2 Transferring passengers

The results for the subset of transferring passengers were obtained using the five behavioural attributes as defined in table 4.1 for the LCA. For the ensemble classifier, the eleven passenger attributes as defined in the same table were used. An overview of all these attributes is given in table 4.9.

[7] Note that this is not the duration of the journey towards the airport, but the time spent at the flight destination before returning 'home'.

Table 4.9: Attributes used in the LCA and classification for transferring passengers

LCA	Classification		
Total transfer time	Total transfer time	Age	Travel purpose
Lounge time	Flight day of week	Destination country	Country of residence
Go to gate time	Travel duration	Origin country	Booking class
Time at gate	Flight frequency	Nationality	

The total number of objects in the subset of transferring passengers is 1,825. The fixed result of the LCA has 1,226 (67%) assigned to class 1, and 599 (33%) to class 2.

4.4.2.1 Classes

The LCA has been performed 50 times, for which the model fit values BIC and AIC are shown in table 4.10. The range of the 95% confidence interval is considerably higher than for the departing passengers subset. However, relative to the mean values these are still relatively small. Hence, the LCA is expected to perform similarly for a single run.

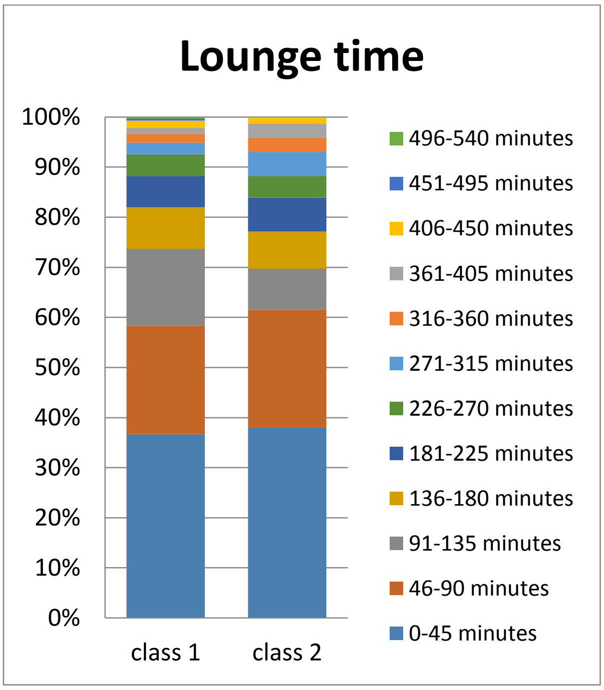
Table 4.10: LCA performance for transferring passengers

Metric	Mean	95% confidence interval
BIC	19491.06	± 78.88
AIC	19210.08	± 78.88

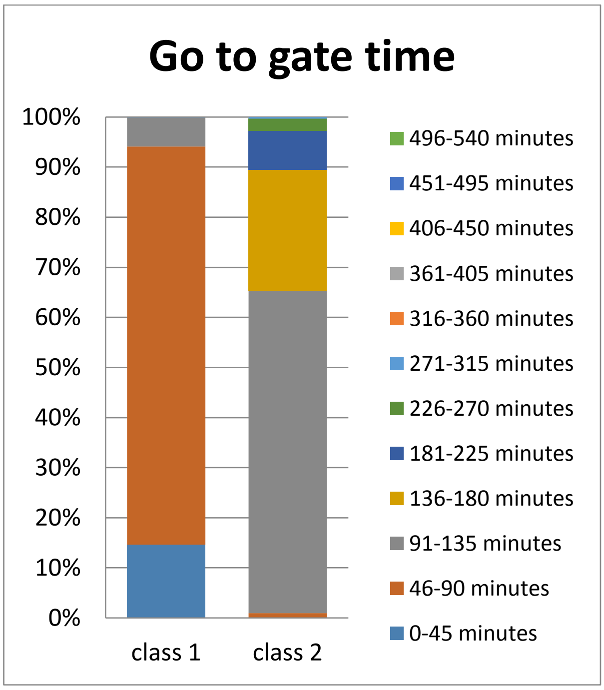
Figure 4.7 shows results of the LCA per class and per attribute outcome-probabilities. Similar to the results of the data set of departing passengers, the predominant difference between the two classes is the time spent at the gate and how far in advance one goes to the gate. In class 1, there is a 94% probability of heading towards the gate less than 90 minutes before flight departure. This is in contrast with class 2, where the probability of heading towards the gate more than 90 minutes in advance is as much as 99%. The probability of heading towards the gate 136 to 180 minutes in advance is even as much as 24%. Not unexpectedly, the gate time attribute shows similar results, where class 1 spends less time at the gate than class 2. However, for this attribute there is more overlap; class 1 has a probability of 62% of spending between 46 and 90 minutes at the gate, whereas class 2 has a probability of 36% for this outcome.

Similar to the results for departing passengers, the difference in the restaurant and shopping attributes is small. Though here the difference in outcome probabilities is even smaller, being virtually the same for both classes. The outcome probabilities of the lounge time attribute is also very similar between the two classes. The probability of spending less than 90 minutes in the lounge is around 60% for both classes. The only difference between the two classes that stands out is the probability of spending between 361 and 405 minutes in the lounge, which is almost double for class 1.

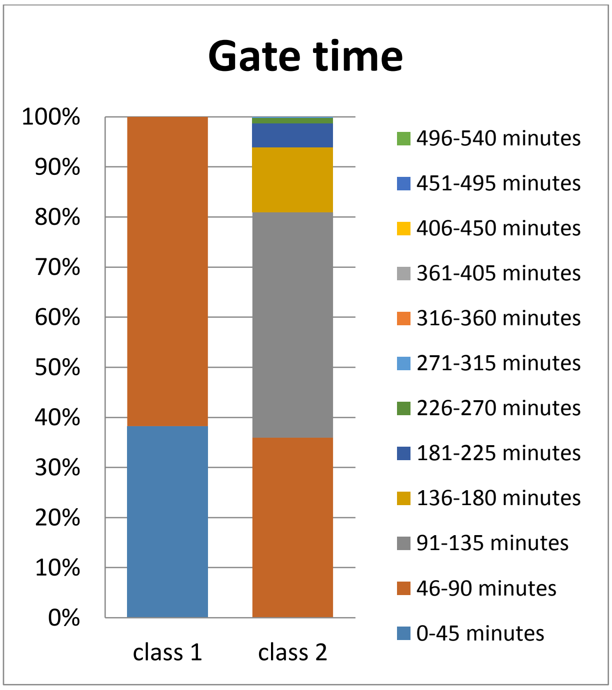
Similarly to the results of departing passengers, characterising the classes would imply class 1 being and 'early-at-the-gate' class and class 2 a 'late-at-the-gate' class.



(a)



(b)



(c)

Figure omitted

(d)

Figure omitted

(e)

Figure 4.7: LCA result for transferring passengers

Similarly to the results of the LCA on the departing passengers, leaving out the time spent at the gate leads to two classes that are primarily distinguished based on the time spent in the lounge and even more so by the shopping and restaurant attributes. Additionally, the performance of the LCA and the classifier is better.

4.4.2.2 Classifier

Table 4.11 shows the mean and 95% confidence interval of the performance metrics for the ensemble classifier. Overall, the performance of the classifier is somewhat better (around 1 percentage point) than for the departing passengers subset.

Table 4.11: Classifier performance for transferring passengers

Set	Metric	Mean	95% confidence interval	
train	AUC	0.9818	±	0.0031
test	AUC	0.6774	±	0.0225
train	Overall accuracy	0.9273	±	0.0074
test	Overall accuracy	0.6733	±	0.0102
train	Average accuracy	0.9273	±	0.0074
test	Average accuracy	0.6733	±	0.0102
train	Macro F1-score	0.9154	±	0.0093
test	Macro F1-score	0.6138	±	0.0169

However, judging from the recall and precision values, presented for both classes in table 4.12, the classifier manages to correctly recognise less than half of the objects that should be in class 2. Also, out of all objects that are classified as class 2, less than half is correct. This may be caused by the fact that for the complete data set, class 2 is significantly smaller than class 1. As a result, the ensemble may be trained better for class 1 than for class 2, resulting in higher recall and precision rates for the former class.

Table 4.12: Per-class recall and precision for transferring passengers

Class	Recall	Precision
Class 1	0.738	0.740
Class 2	0.451	0.449

Figure 4.8 shows the importance of all predictors used for creating the classifier. The predictors are dominated by one attribute: the total transfer time, which has an importance of 40%. Leaving out this attribute, however, significantly reduces the performance of the classifier. In addition to the total transfer time, the day of the week and the travel duration are important predictors for the class.

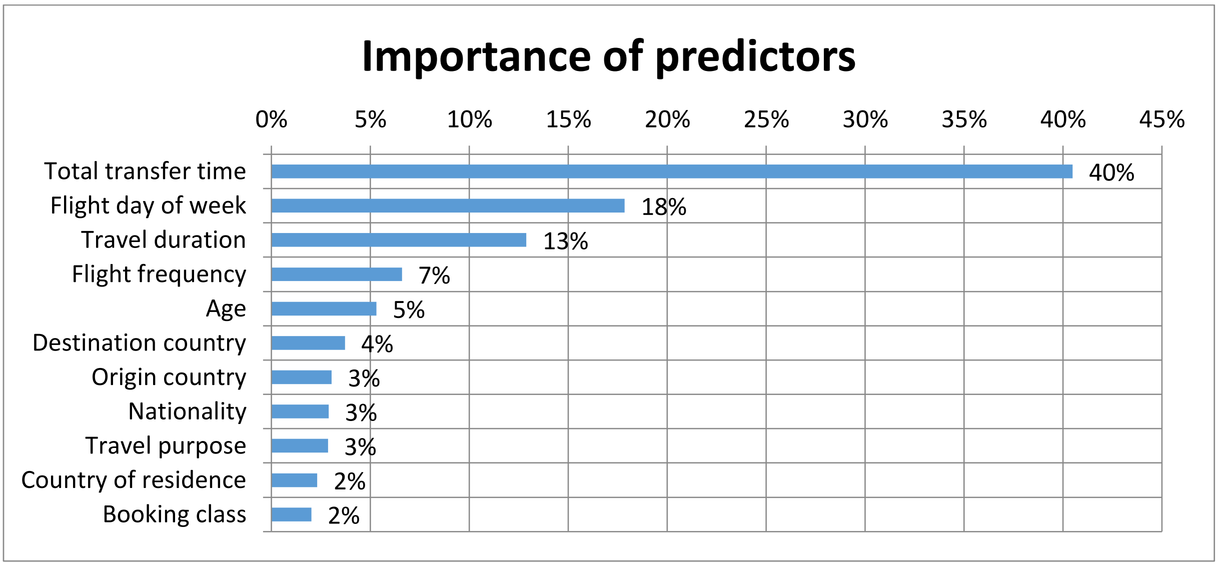


Figure 4.8: Importance of predictors for transferring passengers

4.5 Chapter conclusion

In this chapter, the framework developed in chapter 3 has been applied to a real data set. For this purpose, several possible data sets have been discussed. The Bliptrack and SSBPC data are both based on a form of RF-positioning, which is one of the possible sources of sensor data as introduced in chapter 2. However, both data sets are very limited with respect to the number and type of attributes and therefore not suitable to use with the CC-framework. Combining the two data sets could have yielded a useable data set. However, there is no information present in the data that would reliably allow this.

Ultimately, the CC-framework has been applied to the PASSME data set. This data set contains relatively many attributes. Six of the attributes that have been shown to be related to passenger behaviour in section 2.2.2, are also present in the PASSME data. However, the data set was not specifically collected for the purpose of behavioural classification and is based on survey data. Because of this, the level of detail in the data is somewhat lacking, and some possibly relevant passenger attributes and behavioural attributes are not available. Relations between the variables in the data have been analysed, based on which some conclusions can be drawn:

- Various relations between passenger characteristics and behavioural characteristics are in agreement with the relations as they have been set out earlier in the report:
    - There is a positive correlation between dwell time at the airport and likelihood of shopping and restaurant visit.
    - There is a negative correlation between travel experience and likelihood of shopping and restaurant visit.
    - Travellers with a destination outside the EU spend more time at the airport compared to passenger with a destination within the EU.
- Passengers staying up to about two hours at the airport spend, on average, most of their time at the gate.
- The behavioural attributes regarding time spent in areas of the airport show a high spread. Two of these attributes are normally distributed. There are no apparent classes visible in these attributes.
- Comparing the distributions of behavioural attributes regarding time spent in areas of the airport across the two categories of the ‘shopping’ and ‘restaurant’ attributes, shows that for departing passengers these differ significantly in all cases. For transferring passengers, there is a significant difference for half of the cases. This implies that classes could be found based on this difference.

- Overall, the results of the analysis are very similar for the transferring and departing subsets of the PASSME data.

The analysis of the PASSME data set hence confirms that there are relations between the passenger attributes and behavioural attributes in the data. Moreover, based on the significant differences between the distributions of the behavioural attributes regarding time and the fact that these show a high spread, imply that there are classes to be found in the data. It can be concluded that applying the CC-framework on this data is appropriate.

After the analysis, the CC-framework has been applied to the PASSME data set. Because the CC-framework consists of clustering and classification, there are multiple parameters that can affect the performance of the results of the framework. As such, a grid search was performed in order to find the optimal parameter settings for both subsets regarding the number of classes, the bin size of the numerical attributes, and the maximum tree depth of the decision trees in the ensemble classifier. Based on the results, some conclusions can be drawn which are applicable to both departing and transferring passengers:

- The LCA fit increases as the bin size increases.
- The optimal LCA fit is achieved at around four to five classes.
- The classifier performance generally increases as the number of classes decreases.
- The classifier performance somewhat increases as the bin size increases.
- The effect of maximum tree size on classifier performance is very limited.

For **departing passengers**, the optimal number of classes was found to be two, with a bin size of 60 minutes and a maximum tree depth of 14. The behavioural attributes of the two classes are distinguished by the time that passengers take for going to the gate and sitting at the gate. In addition, a slight difference is visible in restaurant and shopping behaviour and the time spent in the lounge. The two classes of departing passengers have been described as a 'late-at-the-gate' class and an 'early-at-the-gate' class. These two classes do not confirm to the observation from the data analysis that the distributions of the behavioural time attributes differ significantly across the categories of the categorical behavioural attributes. However, leaving the 'gate time' and the 'go to gate time' out of the application of the framework, results in classes that are distinguished by the lounge time, and restaurant and shopping attributes, which does agree with this observation. Nevertheless, the behavioural attributes of the classes that have been found do not contradict. As such, the classes can be considered to be logical. The classification of departing passengers using the CC-framework has yielded an average accuracy of only 64% and an F1-score of 63%.

For **transferring passengers**, the same process has been applied. The results for this subset are very similar to those of the departing passengers: two classes, with a bin size of 45 minutes and a maximum tree depth of 10. Again, the main difference between the two classes is the time spent at the gate and the go to gate time. As such, these classes can also be described as a 'late-at-the-gate' class and an 'early-at-the-gate' class. The late-at-the-gate class spends slightly more time in the lounge. However, the shopping and restaurant behaviour for both classes is almost exactly the same. Furthermore, again similar to the case of departing passengers, leaving the time spent at the gate out of the classification yields two classes that are separated mainly by the difference in time spent at the lounge. Also, in that case there is a significant difference in shopping and restaurant behaviour across the two classes. The classification accuracy on transferring passengers is somewhat higher at 67%, with an F1-score of 61%.

Considering the framework performance requirements set up in section 3.1.4, it can be concluded that the performance of the framework on the PASSME meets the requirements. This means that there are indeed classes in the data, and the performance of classification is higher than random guessing. However, seeing the achieved accuracy and F1-scores, the classification performance cannot be regarded as very high. Several possible causes for the low performance can be considered (and negated):

- The level of detail in the PASSME data is too limited with respect to the number of attributes and the number of response categories for many attributes. Passenger attributes that have been shown to be related to behaviour are not included in the data set. Moreover, apart from restaurant and shopping behaviour, there is no information with respect to the number and duration of activities that passengers have performed in the terminal.
- Binning the numerical behavioural attributes, required by the R package that has been used for clustering, leads to some loss of information. However, the effects of this have been taken into account by applying the framework for many different bin sizes, which resulted in rather large bin sizes of 45 and 60 minutes.
- There are actually no classes present in the data. However, this is quite unlikely based on the results of the data analysis. Moreover, the performance metrics of the LCA indicate a better fit for models with more than one class.

5

Conclusions and Recommendations

At the start of this thesis report, we set out to create a methodology for finding behavioural classes of passengers and classifying passengers to these classes to help predict their behaviour in a behavioural model. Recall the main research question that was introduced there: “How can individual airport passengers be classified according to (visual) sensor-obtained personal characteristics in behavioural classes that can be used to predict passenger behaviour?”. To answer this research question, the subject was divided into three main blocks: Sensing, Processing and Modelling. For each of these blocks, one or more sub-questions were formulated.

In this last chapter, the main research question and sub-questions will be answered. To this end, the chapter is structured as follows. First, section 5.1 expounds the findings of the report. The research questions are then answered in section 5.2. The chapter closes off with section 5.3 by recommending further research regarding behavioural classification.

5.1 Findings

In this section, the findings with respect to the development of this methodology are presented. The findings with respect to Sensing and Modelling are discussed first; these provide the conditions in which the behavioural classification operates. The behavioural classification, part of the Processing block, is discussed after that.

5.1.1 Sensing

With respect to passengers and their behaviour in the terminal, the passenger process has been discussed. This is a rigid, linear process. It can be divided into a few phases, separated by mandatory processes. During these phases, passengers are able to perform discretionary activities. Although there is some overlap, the process is fairly different for departing, arriving, and transferring passengers. Consequently, these passenger types should be treated separately in a behavioural classification.

To describe the characteristics of passengers and their behaviour, two definitions for characteristics have been set:

- **Behavioural characteristics** are the characteristics that are to be predicted by the behavioural classification. Each behavioural class has its specific behavioural characteristics. These behavioural characteristics form the input for a behavioural model.
- **Passenger characteristics** are the characteristics based on which passengers are classified into a behavioural class. These can be visible or non-visible characteristics, such as age, sex, flight destination and travel class.

There are several proven relations between passenger characteristics and behaviour according to literature. These are: age, airline type, amount of carry-on baggage, check-in method, education level, gender, group composition, income, place of residence/travel destination/travelling company, total time spent at the airport, travel class, travel destination, travel experience, and travel purpose. These characteristics have been shown to be related to one or more of the following aspects of behaviour: activity set, likelihood of going shopping or dinning, and the time spent in the various phases of the passenger process.

For nearly all of these possible characteristics, a possible source of data is (theoretically) available, either in the form of visual sensors, RF-positioning, airport databases, et cetera. However, though many possible sources of data are already available, or could be implemented, there are two main challenges that should be tackled. First, as the data come from different sources, there is no universal identifier describing to which passenger the data applies. Fusing these data from these sources into one database is a challenge, though it would result in a highly detailed data set on an individual level. Second, even though such a data set could be anonymised so it does not contain personal details, privacy remains a concern.

5.1.2 Modelling

As the final goal of the behavioural classification is to create a more accurate representation of passenger behaviour in a future PASSME PDF system, behavioural models and theory with respect to passenger behaviour have been discussed. Broadly, the behaviour of a passenger can be characterised as a process with three levels of decision-making: strategic, tactical and operational. Although most behavioural models focus on the tactical and operational level, an ideal behavioural model would accurately model all levels of behaviour. Hence, this implies that for a good behavioural classification, behavioural characteristics of passengers should be available on all levels. In order to effectively tune the behavioural classification of this thesis to a behavioural model, it should be known what behavioural parameters are present the model that the classification forms the input to.

5.1.3 Processing

For the Processing block, the clustering and classification (CC) framework for passengers has been introduced. The framework performs two main tasks: clustering and classification. For both tasks, various methods have been considered. For clustering, Latent Class Analysis was chosen because it has been shown to yield good results, is able to handle mixed-type data, and it provides good performance metrics because it is model-based. For classification, a boosted ensemble of decision trees is used. These have been chosen because they are relatively quick to compute, offer a transparent classifier, and support mixed-type data.

To test the implemented CC-framework, the PASSME data was used. These data comprise of around 4,000 survey results of departing and transferring passengers at an airport. The set contains a limited number of attributes describing behaviour, which are not very detailed. The statistical relations between the attributes in the data were assessed. Based on this, it was determined that there are several significant relations between the attributes in the data set, and that is likely that behavioural classes could be found in the data.

From the application of the CC-framework on the PASSME data set, it could be concluded that the performance of the classifier is optimal for a low number of classes, and a large bin size for the continuous time attributes in the data. This resulted in a classification with two classes, both for the data set of departing passengers as for the data set of transferring passengers. It has been shown that the main difference between the two classes is in both cases based mainly on the attributes regarding the time spent at the gate and the time of going to the gate. Additionally the performance of the classifier is quite low for both cases; average accuracies are around 65%. A detailed overview of the classification performance is given in table 5.1

Table 5.1: Overview of behavioural classification results

Metric	Departing passengers	Transferring passengers
AUC	0.6883	0.6774
Overall accuracy	0.6363	0.6733
Average accuracy	0.6363	0.6733
Macro F1-score	0.6250	0.6138

5.2 Conclusions

The research questions have been formulated per thesis block. One question each has been formulated for Sensing and Modelling. Three questions were formulated for Processing. These sub-questions will be answered here. After this, the main research question will be answered.

5.2.1 Sensing sub-question

Which passenger characteristics that can be used for behavioural classification can be obtained from sensors and information systems in an airport terminal?

In an airport, there are quite a few possible sources of data that can theoretically be used to collect passenger characteristics. Viable options for an airport terminal are RF-positioning, video analysis, 3D vision, airport apps, the airport database, and the airlines databases. Looking at the example of AAS, there are already several systems in place that collect interesting types of data, such as Bliptrack and SSBPC.

From all these possible sources of data, many different passenger characteristics can be collected. The airport and airline data can mainly provide information about the passenger's trip, i.e. origin, destination, flight number, et cetera. Video analysis and an airport app could mainly provide information about the personal properties of a passenger, i.e., age, mood, sex, et cetera. Lastly, RF-positioning can provide information about the location and activities of the passenger.

5.2.2 Modelling sub-question

How could behavioural classes form the input to behavioural models such as in a passenger demand forecast system?

Behaviour of passengers consists of three levels: strategic, tactical, and operational. Each of these levels describes a part of passenger behaviour. An ideal model would simulate all three levels. However, most behavioural models model one or two levels. The other levels are then exogenous to the model and are required as input, depending on the specifics of the model. Moreover, there are multiple types of models: macroscopic, mesoscopic and microscopic. Each of these models requires different inputs. Evidently, there is no universal answer as to what input behavioural models require.

However, on a more general level, it can be noted how a behavioural classification as created in this report can interface with behavioural models. The behavioural attributes of the behavioural classes following from behavioural classification should resemble the input attributes of the behavioural model. This means that the behavioural classes contain parameter values that can be used in behavioural models.

5.2.3 Processing sub-questions

Processing sub-question 1: *Which behavioural classes can be used to predict passenger behaviour?*

Following the results from the CC-framework applied to the PASSME data set, two behavioural classes have been found by optimizing the CC-framework performance based on classifier accuracy. The two behavioural classes differ mainly with respect to the time passengers spend at the gate, and the time remaining until flight departure when passengers decide to go to the gate. These behavioural classes have therefore been interpreted as a 'late-at-the-gate' class and an 'early-at-the-gate' class. These classes are valid both for departing, as well as transferring passengers.

However, these classes are evidently only valid for this specific data set. Using the framework on other data with other or more attributes can lead to different classes. Additionally, although this did not appear from the PASSME data, classes could vary over time for different days of the week, seasons, holidays, et cetera.

Processing sub-question 2: *How can passengers be classified to these behavioural classes?*

Based on the CC-framework, passengers can be classified according to their passenger characteristics. To do this, a classifier is needed that is trained based the clustered data set. The classification algorithm assesses the passenger attributes and the assigned class of each passenger in the training data. Based on this, the classifier is constructed.

The classification algorithm used for the CC-framework is the SAMME algorithm. This algorithm is essentially a multiclass adaptation of the well-known and well-performing AdaBoost algorithm. It uses a boosted ensemble of decision trees. Tests using the PASSME data set have indicated that this methodology leads to an F1-score, indicating the performance of the classifier with respect to precision and recall, of about 63% for departing passengers, and 61% for transferring passengers.

Processing sub-question 3: *When is the result of classification satisfactory?*

The performance of classification can be expressed using various performance metrics. In this report, AUC, overall accuracy, average accuracy, precision, recall, and F1-score have been used. None of these metrics are all-encompassing with respect to representing classification performance as a single number. Additionally, the classification performance that is to be expected also greatly depends on the type of data used for classification. Consequently, it is argued in this report that the classification should at least exceed the performance of randomly objects assigning to classes. Such performance is achieved at an AUC equal to 0.5; a classifier with an AUC above 0.5 is hence better than random assignment.

Relating this to the performance of the CC-framework on the PASSME data, it can be noted that the performance is in both cases higher than this threshold. As such, according to the definition, such a classification can be regarded as satisfactory.

5.2.4 Main research question

Main research question: *How can individual airport passengers be classified according to (visual) sensor-obtained personal characteristics in behavioural classes that can be used to predict passenger behaviour?*

Passengers can be classified into behavioural classes using the Clustering and Classification framework developed in this report. In short, the steps associated with this are as follows:

- From various (sensor) data sources, passenger characteristics and behavioural characteristics of passengers are collected on an individual level.
- Aggregating these individual observations into a larger data set yields a training data set that can be used to train the CC-framework:
    - The first part of the CC-framework, clustering, is done using LCA. This LCA is performed only on the behavioural characteristics in the data. This yields the behavioural classes. The training data set now also contains a class number for each individual in the data set, apart from the behavioural characteristics and passenger characteristics.
    - The second part of the CC-framework, classification, is done using an ensemble classifier that uses decision trees, based on the SAMME algorithm. The classifier is built using only the passenger attributes and the behavioural class of each individual in the training data set.
- New cases of passengers that were not part of the training data set can now be classified based solely on their passenger attributes. The behavioural attributes of the behavioural

class they are classified to are used in the behavioural model that will predict their behaviour.
- When necessary, the CC-framework can be retrained in order to create new behavioural classes. This could be the case when different classifiers are required for weekends and weekdays, different seasons, months, et cetera.

The average per-class accuracy and macro F1-score that were achieved by applying the framework to the PASSME data are around 60% to 65%. Though satisfying the requirement of performing better than random assignment (the AUC is above 0.5 in all cases), the classification performance is not particularly high. However, it should be kept in mind that the data to which the framework has been applied, was limited; not all desired attributes are available in the data, and the level of detail is low due to being based on surveys that were not collected for the purpose. Therefore, it can be concluded that the CC-framework yields promising results that allow for predicting behaviour. Recommendations regarding further development and improvement of the framework will be given in the next section.

5.3 Recommendations and further research

Based on the aforementioned findings and conclusions, several recommendations for practice and for further research can be made:

- Application of the CC-framework on the PASSME data set has given promising results, given the aforementioned shortcomings of the data. For further testing of the framework, it is recommended to use a more adequate set of data with more and more detailed attributes, at least satisfying the list of passenger attributes given in section 2.2.2, though preferably satisfying the overview of characteristics given in Appendix A.
- The possible interaction of behavioural classification with a behavioural model in a PASSME PDF has been discussed on a general level. Further research with respect to this interaction should be conducted once the PDF is further developed.
- The CC-framework has been implemented using the R programming language. Although this language is very apt for the present initial research, it is limited in terms of performance and flexibility. Especially the classification can take a very long time as the model complexity increases. Should the framework be further developed, it is advisable to switch to another programming language.
- The implemented LCA does not support continuous variables. Because of this, continuous variables in the PASSME data had to be discretised. Although the classification result has also been optimised based on the best bin size, discretisation does inherently cause some loss of information. Hence, a future improvement of the implementation should support continuous variables.
- The optimisation of the parameters of clustering and classification requires human interpretation of various performance metrics. This makes it difficult to recognise the best result. The creation of one indicator or objective function based on which the best solution for the CC-framework can (automatically) be found, can help achieving the best framework performance.
- For the PASSME data set, it has been shown that leaving out some attributes out of the behavioural classes, results in different behavioural classes. More research can be done with respect to the effects of this. Creating behavioural classes for subsets of behavioural attributes could possibly be useful.

This page intentionally left blank.

Bibliography

Abdelghany, A., Abdelghany, K., & Mahmassani, H. (2016). A hybrid simulation-assignment modeling framework for crowd dynamics in large-scale pedestrian facilities. *Transportation Research Part A: Policy and Practice, 86*, 159-176. doi:10.1016/j.tra.2016.02.011

Abonyi, J., & Feil, B. (2007). *Cluster analysis for data mining and system identification*: Springer Science & Business Media.

Airports Council International (ACI). (4 April 2016). ACI releases preliminary world airport traffic rankings. Retrieved from http://www.aci.aero/News/Releases/Most-Recent/2016/04/04/ACI-releases-preliminary-world-airport-traffic-rankings

Akaike, H. (1998). Information Theory and an Extension of the Maximum Likelihood Principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 199-213). New York, NY: Springer New York.

Alfaro, E., Gámez, M., & Garcia, N. (2013). Adabag: An R package for classification with boosting and bagging. *Journal of Statistical Software, 54*(2), 1-35.

Ali, S., Nishino, K., Manocha, D., & Shah, M. (2013). Modeling, Simulation and Visual Analysis of Crowds: A Multidisciplinary Perspective *Modeling, Simulation and Visual Analysis of Crowds* (pp. 1-19): Springer.

AT5. (22 April 2016). Extra personeel Schiphol tegen lange wachtrijen. Retrieved from http://www.nu.nl/amsterdam/4251383/extra-personeel-schiphol-lange-wachtrijen.html

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition, 30*(7), 1145-1159. doi:http://dx.doi.org/10.1016/S0031-3203(96)00142-2

Castillo-Manzano, J. I., & López-Valpuesta, L. (2013). Analysing passenger behaviour towards the catering industry: Implications for airport management. *International Journal of Hospitality Management, 35*, 258-260. doi:http://dx.doi.org/10.1016/j.ijhm.2013.06.010

Cheung, Y.-M., & Jia, H. (2013). Categorical-and-numerical-attribute data clustering based on a unified similarity metric without knowing cluster number. *Pattern Recognition, 46*(8), 2228-2238.

Cristiani, E., Piccoli, B., & Tosin, A. (2014). An Introduction to the Modeling of Crowd Dynamics *Multiscale Modeling of Pedestrian Dynamics* (pp. 3-27). Cham: Springer International Publishing.

Daamen, W. (2004). *Modelling passenger flows in public transport facilities*: TU Delft, Delft University of Technology.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1-38.

Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning, 40*(2), 139-157.

Duives, D. C., Daamen, W., & Hoogendoorn, S. P. (2013). State-of-the-art crowd motion simulation models. *Transportation Research Part C: Emerging Technologies, 37*, 193-209. doi:10.1016/j.trc.2013.02.005

Dziak, J. J., Coffman, D. L., Lanza, S. T., & Li, R. (2012). Sensitivity and specificity of information criteria. *The Methodology Center and Department of Statistics, Penn State, The Pennsylvania State University*, 1-10.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874.

Freund, Y., & Schapire, R. E. (1996). *Experiments with a new boosting algorithm.* Paper presented at the Icml.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 857-871.

Grosmann, R. J. F. (2016). *D4.1 Operational Concept of Passenger Demand Forecast System*. Retrieved from http://www.passme.eu/results-to-date

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*: Elsevier.

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine learning, 45*(2), 171-186.

Harle, R. (2013). A survey of indoor inertial positioning systems for pedestrians. *IEEE Communications Surveys & Tutorials, 15*(3), 1281-1293.

He, X., Cai, D., & Niyogi, P. (2005). *Laplacian score for feature selection.* Paper presented at the Advances in neural information processing systems.
Hoogendoorn, S. P. (2001). *Normative pedestrian flow behavior theory and applications* (0920-0592). Retrieved from
Hoogendoorn, S. P., & Bovy, P. H. L. (2004). Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological, 38*(2), 169-190. doi:10.1016/s0191-2615(03)00007-9
Hu, W., Tan, T., Wang, L., & Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34*(3), 334-352.
Kalakou, S., & Moura, F. (2015). Modelling Passengers' Activity Choice in Airport Terminal before the Security Checkpoint: The Case of Portela Airport in Lisbon. *Transportation Research Procedia, 10*, 881-890. doi:10.1016/j.trpro.2015.09.041
Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques.
Kountouriotis, V., Thomopoulos, S. C. A., & Papelis, Y. (2014). An agent-based crowd behaviour model for real time crowd behaviour simulation. *Pattern Recognition Letters, 44*, 30-38. doi:10.1016/j.patrec.2013.10.024
Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. from University of California, School of Information and Computer Science
Lin, Y.-H., & Chen, C.-F. (2013). Passengers' shopping motivations and commercial activities at airports–The moderating effects of time pressure and impulse buying tendency. *Tourism Management, 36*, 426-434.
Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software, 42*(10), 1-29.
Liu, H., Darabi, H., Banerjee, P., & Liu, J. (2007). Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 37*(6), 1067-1080.
Liu, X., Usher, J. M., & Strawderman, L. (2014). An analysis of activity scheduling behavior of airport travelers. *Computers & Industrial Engineering, 74*, 208-218. doi:10.1016/j.cie.2014.05.016
Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography, 17*(2), 145-151. doi:10.1111/j.1466-8238.2007.00358.x
Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research, 20*(1), 36-43.
Millonig, A., Brändle, N., Ray, M., Bauer, D., & van der Spek, S. (2009). *Pedestrian Behaviour Monitoring: Methods and Experiences.* Paper presented at the BMI Book.
Molin, E., Mokhtarian, P., & Kroesen, M. (2016). Multimodal travel groups and attitudes: A latent class cluster analysis of Dutch travelers. *Transportation Research Part A: Policy and Practice, 83*, 14-29.
NU.nl. (20 April 2016). Schiphol adviseert reizigers op tijd te komen rondom meivakantie. Retrieved from http://www.nu.nl/lifestyle/4249923/schiphol-adviseert-reizigers-tijd-komen-rondom-meivakantie.html
Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research, 11*, 169-198.
Otsason, V., Varshavsky, A., LaMarca, A., & De Lara, E. (2005). *Accurate GSM indoor localization.* Paper presented at the International conference on ubiquitous computing.
Ouyang, W., & Wang, X. (2012). *A discriminative deep model for pedestrian detection with occlusion handling.* Paper presented at the Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.
PASSME. (n.d.-a). About the PASSME project. Retrieved from http://www.passme.eu/about-passme
PASSME. (n.d.-b). PASSME and the Future of Passenger Demand Forecasting. Retrieved from http://www.passme.eu/related-news/163-passme-and-the-future-of-passenger-demand-forecasting

Phua, P., Page, B., & Bogomolova, S. (2015). Validating Bluetooth logging as metric for shopper behaviour studies. *Journal of Retailing and Consumer Services, 22*, 158-163. doi:http://dx.doi.org/10.1016/j.jretconser.2014.10.009

Poushter, J. (2016). Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. *Pew Research Center: Global Attitudes & Trends*.

R Core Team. (2015). R: A Language and Environment for Statistical Computing (Version 3.2.3 (2015-12-10)). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics, 12*(1), 1.

Schadschneider, A., Klüpfel, H., Kretz, T., Rogsch, C., & Seyfried, A. (2009). Fundamentals of Pedestrian and Evacuation Dynamics. 124-154. doi:10.4018/978-1-60566-226-8.ch006

Schiphol Group (Producer). (2016). Passenger Journey 2016 (Schiphol internal document).

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics, 6*(2), 461-464.

Sightcorp. (n.d.). Crowdsight SDK. Retrieved from http://sightcorp.com/crowdsight/

Silveira Jacques Junior, J., Musse, S., & Jung, C. (2010). Crowd Analysis Using Computer Vision Techniques. *IEEE Signal Processing Magazine*. doi:10.1109/msp.2010.937394

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427-437.

Tam, M. L., Lam, W. H., & Lo, H. P. (2008). Modeling air passenger travel behavior on airport ground access mode choices. *Transportmetrica, 4*(2), 135-153.

Torres, E., Domínguez, J. S., Valdés, L., & Aza, R. (2005). Passenger waiting time in an airport and expenditure carried out in the commercial area. *Journal of Air Transport Management, 11*(6), 363-367. doi:http://dx.doi.org/10.1016/j.jairtraman.2005.04.001

Tran, K. N., Gala, A., Kakadiaris, I. A., & Shah, S. K. (2014). Activity analysis in crowded environments using social cues for group discovery and human interaction modeling. *Pattern Recognition Letters, 44*, 49-57. doi:10.1016/j.patrec.2013.09.015

Vermunt, J. K., & Magidson, J. (2004). Latent class analysis. *The sage encyclopedia of social sciences research methods*, 549-553.

Vorst, P., Sommer, J., Hoene, C., Schneider, P., Weiss, C., Schairer, T., . . . Carle, G. (2008). *Indoor positioning via three different RF technologies.* Paper presented at the RFID Systems and Technologies (RFID SysTech), 2008 4th European Workshop on.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on neural networks, 16*(3), 645-678.

Zaki, M. H., & Sayed, T. (2014). Using automated walking gait analysis for the identification of pedestrian attributes. *Transportation Research Part C: Emerging Technologies, 48*, 16-36. doi:10.1016/j.trc.2014.08.004

Zhan, B., Monekosso, D. N., Remagnino, P., Velastin, S. A., & Xu, L.-Q. (2008). Crowd analysis: a survey. *Machine Vision and Applications, 19*(5-6), 345-357. doi:10.1007/s00138-008-0132-4

Zhao, Q. (2012). Cluster validity in clustering methods. *Publications of the University of Eastern Finland Dissertations in Forestry and Natural Sciences, 77*.

Zhu, J., Rosset, S., Zou, H., & Hastie, T. (2006). Multi-class adaboost. *Ann Arbor, 1001*(48109), 1612.

Zhu, S., Ji, X., Xu, W., & Gong, Y. (2005). *Multi-labelled classification using maximum entropy method.* Paper presented at the Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval.

This page intentionally left blank.

Appendix A
Overview of characteristics and data sources

					Available technology (green)								L	Linked (to identified/tagged passenger)		
					Possible technology (yellow)								U	Unlinked		
					Future technology (red)											
No.	**Type**	**Category**	**Characteristic**	**Values**	**Airport Database**	**Airline Database**	**Touch point**	**Bluetooth**	**Wi-Fi**	**Beacons**	**Video analysis**	**3D Vision**	**Mobile app**	**Comment - available technology**	**Comment - possible technology**	**Comment - future technology**
1	Passenger	Personal	Age	Number		L					L		L	When booking a ticket, it depends on the destination if one has to provide age information. However, in some cases this may still be available in the airline database, for example due to a frequent flyer programme.	Video analysis software, such as Crowdsight (developing) can estimate age. Literature on categorizing age based on walking gait analysis is also available.	In the future, retrieving the age of a passenger may be possible through an app, such as the PASSME app. For this, it is required that the app asks the traveller to enter his/her date of birth when he/she starts using the app.
2	Passenger	Personal	Physique	Length/footprint							L			There is currently no information available about a passenger's physique.	Pedestrians can currently be detected in video, but height and width detection is inaccurate. Occlusion and camera setup is a problem.	Future improvements in computer vision can increase the accuracy of height/width determination.

3	Passenger	Personal	Gender	M/F		L	L				L		L	For some destinations, passengers have to provide their gender, making it available in an airline database. Non-Schengen passengers have to show their identification at border control. However, data from this could not be used due to privacy regulations.	Crowdsight and research in literature show that it is possible to determine gender based on video analysis methods, for example walking gait analysis.	A future PASSME app can ask the user to provide personal details such as gender.
4	Passenger	Personal	Nationality	Country		L	L						L	For some destinations, passengers have to provide their nationality, making it available in an airline database. Non-Schengen passengers have to show their identification at border control. However, data from this could not be used due to privacy regulations.		A future PASSME app can ask the user to provide personal details such as nationality. Alternatively, the app can guess the nationality of the passengers based on the locale setting of the device.
5	Passenger	Personal	Ethnicity	Race (Asian, African, Caucasian, Hispanic, Etc.)							L			There is currently no information available with respect to the ethnicity of a passenger.	Video analysis software, such as Crowdsight can estimate ethnicity based on computer vision.	Future developments may improve the accuracy of ethnicity estimation.

6	Passenger	Personal	Group composition	Single/Couple/Family/Friends/Tour group		L					U			A ticket can be part of one booking containing multiple tickets. This can be an indication that a passenger travels in a group and should be available from the airline database. Based on the characteristics of the group members, an estimation about the group type can be made.		Video analysis techniques can identify groups in an image. The type of group may be estimated based on the analysis of interactions within the group. However, this not possible with current techniques.
7	Passenger	Personal	Group size	Number		L					U			A ticket can be part of one booking containing multiple tickets. This can be an indication that a passenger travels in a group and should be available from the airline database.	Video analysis techniques can identify groups in an image. To be able to link this to the passengers in the image, these passengers have to be identified. This also brings occlusion problems.	Further technological developments can increase the performance of video-based group detection.
8	Passenger	Personal	Role	Leader/Follower/Wanderer							U					Leader/Follower behaviour is very complicated and not directly observable. Future video analysis may provide information about this. Again, observed behaviour has to be linked to an identified passenger

9	Passenger	Personal	Carry-on baggage	Backpack/Suitcase/Trolley/None							U	U		Neither the airport, nor the airline has any knowledge about the amount of carry-on baggage carried.	Video analysis techniques can identify distinct carry-on baggage such as a small trolley.	Further developed video analysis and 3D vision techniques may be able to distinguish more types of carry-on luggage.
10	Passenger	Personal	Holding baggage	Backpack/Suitcase/Trolley/None	L	L					U	U		The airline registers the amount of checked-in baggage, its weight and whether or not it is odd-size. This is linked to a specific passenger name. Additionally, the airport processes baggage and scans the attached baggage labels, which includes the passenger's name.		Upon terminal entrance (hence before check-in/baggage drop-off), passengers can be detected by cameras and/or 3D vision techniques, which can distinguish the amount and type of carried hold baggage.
11	Passenger	Personal	Mood	Attribute values (happy, surprised, anger, disgusted, afraid, sad)							L		L	There is currently no information available about the mood of passengers, except, for example, observations by floor managers.	Video analysis tool Crowdsight is able to represent a person's mood based on six basic mood indicators. This requires a good coverage of cameras in order to mitigate occlusion effects.	In the PASSME app, passengers will be able to give feedback about their mood. However, people only tend to report negative experiences. Nevertheless this could imply that passengers using the app that did not report a bad mood can be considered to be in a good mood.

12	Behavioural	Personal	Walking speed	Number				U	U	U	U	U		Based on existing tracking information from systems such as Bliptrack (RF tracking) a basic walking speed estimation can be derived. RF tracking is based on MAC address, which has to be linked to an identified passenger.	Walking speed is a relatively basic characteristic that can be observed by video analysis/3D vision. Once again, the main challenge here is to link this observation to a specifically identified passenger.	
13	*Passenger*	*Personal*	*Intelligence/education*	*Category*										*Except by means of questionnaires, it is not possible to observe one's intelligence/education level.*		
14	*Passenger*	*Personal*	*Wealth*	*Wealthy/Poor*										*Not directly available in any case, but may be estimated based on e.g. travel destination, clothing, and location in terminal.*		
15	Behavioural	Personal	Mobility	Able-bodied/partially disabled/disabled	L	L					U			People who need assistance for any reason can report so to the airline or airport, in which respective database this will be registered.	Video analysis techniques can distinguish objects such as walking canes or wheelchairs. Occlusion and detection rate are limiting factors.	
16	Passenger	Personal	Experience	Category (novice-experienced)		L					U			Passengers registered in an airline loyalty programme will generally be more experienced travellers. This data is available from the airline.		Further advancements in behaviour analysis based on video may provide more information about the travel experience of a passenger.

17	Behavioural	Personal	Assertiveness (walking behaviour)	Category (Not assertive - very assertive)				U	U	U	U					Assertiveness is not directly observable, but is rather a result of interaction with other pedestrians and the environment. This may be observable based on the analysis of walking trajectories.
18	Behavioural	Personal	Obedience	Rule-abiding/rule-ignoring				U	U	U	U					Obedience is not directly observable, but may be available based on analysis of expected behaviour and actual behaviour.
19	Behavioural	Personal	Location	X, Y, Z			U	U	U	U	U	U		Current RF technologies such as Bliptrack register the location of a passenger based on MAC address, which has to be linked to an individual passenger.	Video analysis and 3D vision can be used to detect and track passengers. The observed passenger has to be identified.	
20	Behavioural	Personal	Traversed path	X, Y, Z, t			U	U	U	U	U	U		Analysis of location over time.	Analysis of location over time.	Analysis of location over time.
21	Passenger	Personal	Passenger/Non-passenger	Category (Passenger/Non-Passenger/Staff)			U				U			Based on touchpoints such as check-in and boarding pass scans, passengers can be distinguished from non-passengers and staff.		More advanced video analytics may be able to identify staff from passengers based on attributes such as tools or clothing.

22	Passenger	Personal	Trip goal	Category (Business/Leisure/Health /Family visit/Friends visit/Education)									L	Trip goal is currently not known, but may be derived based on fare class.		The future PASSME app may inquire the user about his or her trip purpose.
23	Passenger	Personal	Part of trip	Transferring/Originating/ Destination			U	U	U	U	U	U	L	In case of one boarding pass for multiple flights, data from touchpoints can help determine if a passenger is a transferring.	Location analysis can help determine what part of the trip a passenger is in. For example, a passenger first observed near a gate and then later for some time in a lounge is likely to be a transferring passenger.	If a passengers enters his itinerary in the app, the app can determine the part of the trip the passenger is in.
24	Passenger	Personal	Flight number	Flight number		L	U				U	U		The airline database contains the flight number based on the passenger's name. Touchpoint data can also provide the flight number, but is not linked to a specific passenger.	Passengers observed at a gate can be assumed to be boarding the flight from that gate.	
25	Passenger	Personal	Passenger appearance (departing from airport)	Time		L	U	U	U	U	U			In case a passenger has to check-in or drop off baggage, the time of appearance can be estimated. Additionally, RF observations at the terminal entrance can determine the time of appearance.	Video analysis near the terminal entrance can determine the time a passenger enters the terminal.	

26	Passenger	Personal	Number of connections remaining	Number		L							L	The airline database contains the full passenger itinerary.		If a passengers enters his itinerary in the app, the app can determine the number of connections remaining.
27	Behavioural	Process	Activity	Activity			U	U	U	U	U	U		Analysis of location can determine the activity a passenger is engaged in.	Analysis of location can determine the activity a passenger is engaged in.	
28	Behavioural	Process	Time spent during each step in the (mandatory) pax process	Number of minutes			U	U	U	U	U	U		Analysis of location over time.	Analysis of location over time.	
29	Passenger	Trip	Origin	Airport (Schengen/Non-Schengen/ICA/100%)		L							L	The airline database contains the full passenger itinerary.		If a passengers enters his itinerary in the app, the app can determine the origin.
30	Passenger	Trip	Destination	Airport (Schengen/Non-Schengen/ICA/100%)		L							L	The airline database contains the full passenger itinerary.		If a passengers enters his itinerary in the app, the app can determine the destination.
31	Passenger	Trip	Hall	Hall number	L	L								Information based on passenger flight number.		
32	Passenger	Trip	Reclaim	Reclaim number	L	L								Information based on passenger flight number.		

33	Passenger	Trip	Aircraft type	Aircraft	L	L								Information based on passenger flight number.		
34	Passenger	Trip	Number of pax on flight	Number		L								Information based on passenger flight number.		
35	Passenger	Trip	Number of transfer pax on flight	Number		L								Information based on passenger flight number.		
36	Passenger	Trip	Gate	Gate number		L								Information based on passenger flight number.		
37	Passenger	Trip	Airline	Airline name		L								Information based on passenger flight number.		

This page intentionally left blank.

Appendix B
Overview of conventional clustering algorithms

Conventional clustering performance metrics

Extrinsic indices

Extrinsic indices compare the achieved clustering to the ground truth. A possible index is the correlation. Both for the ground truth and the cluster result, a matrix is constructed with one row and one column for each object. For each cell i, j, object i is compared to object j. A value of 1 is given if the pair of objects belongs to the same cluster. A value of 0 indicates the converse. Combining the matrices of the ground truth and the cluster result yields the total number of errors. As the matrix is symmetrical, only half of the matrix should be regarded. The correlation is then calculated according to:

$$correlation = \frac{errors}{\frac{n(n-1)}{2}} \tag{0.1}$$

Another extrinsic index is presented in the work of He, Cai, and Niyogi (2005) and is calculated according to:

$$ACC = \frac{\sum_{i=1}^{N} \delta(c_i, map(l_i))}{N} \tag{0.2}$$

Here, N is the number of objects in the data set. The function $map(l_i)$ maps the cluster labels l_i to the corresponding ground truth labels using a mapping algorithm such as the Kuhn-Munkres algorithm. The delta function assumes a value of 1 if the ground truth label $c_i = map(l_i)$. Otherwise, the function assumes a value of 0.

Intrinsic indices

Intrinsic indices are not based on a ground truth but rather assess the cluster validity based on the data itself. There are various intrinsic indices available. Due to the imperfectness of each one, it is advised to use multiple indices to assess validity (Abonyi & Feil, 2007). Some examples of intrinsic indices include the silhouette coefficient (Han et al., 2011), partition coefficient and classification entropy (Abonyi & Feil, 2007; Zhao, 2012).

Other clustering algorithms explained

In addition to the latent class analysis algorithm that has been implemented in the report, some alternative, 'regular' clustering algorithms have been tested. Although these algorithms did not yield very good results on the data set, the choice for these algorithms and a general description of their workings is discussed below.

Because of size of the data set that is expected to be used in an operational situation, it was decided to use non-hierarchical algorithms. Additionally, because of the fact that passenger behaviour can be regarded as a natural phenomenon and the intuitive notion of travellers being better represented by a combination of multiple clusters, such as explained in section 3.3.1.2, it was chosen to implement a fuzzy algorithm. Two fuzzy algorithms with a modified distance measure to account for mixed variables were implemented: the expectation-maximisation (EM) and the fuzzy c-means (FCM) algorithms. These two algorithms are primarily designed to be used with numerical attributes. However, these algorithms can be made suitable for use with categorical attributes or mixed data sets. This has been done by using an alternative distance metric, the Gower distance, or by transforming categorical attributes into multiple binary values and treating them as numerical values.

In addition to the two fuzzy algorithms, one non-fuzzy algorithm specifically designed for data sets with mixed attributes was used: the iterative clustering learning based on object-cluster similarity metric (OCIL) algorithm (Cheung & Jia, 2013). The algorithms have been implemented in MATLAB. To assess their performance, various synthetic data sets were produced. Additionally, data sets from the UCI repository were used. The following algorithms were implemented: Expectation-Maximisation (EM), Fuzzy c-means (FCM) and the OCIL algorithm (Cheung & Jia, 2013).

Expectation-maximisation algorithm (EM)

The EM algorithm is a relatively simple algorithm that can perform fuzzy clustering. The implemented EM algorithm is described in the work of Han et al. (2011). The algorithm was implemented using the Gower distance measure, with an option to use Euclidean distance. The algorithm consists of the following steps:

Input: data set, desired number of clusters K

1. For K clusters, initiate K random objects as cluster prototypes
2. For each object, calculate Gower distance to each cluster
3. Based on the Gower distance, proportionally assign membership values for each object to each cluster
4. Recalculate cluster prototypes based on the membership values w_{i,c_j} and the attribute value o_i for all clusters and all objects, using (0.3).

$$c_j = \frac{\sum_{i=1}^{N} w_{i,c_j}^2 o_i}{\sum_{i=1}^{N} w_{i,c_j}^2}, \forall j \in K \quad (0.3)$$

5. Return to step 2 and repeat as desired

Fuzzy c-means algorithm (FCM)

The fuzzy c-means is arguably the best-known fuzzy clustering algorithm. The algorithm has been implemented using the Gower distance measure, with an option to use Euclidean distance.

Input : data set, desired number of clusters C, fuzziness parameter $m > 1$

1. For C clusters, initiate C random objects as cluster prototypes
2. Calculate membership values u_{ij} for each object i with respect to each cluster j according to (0.4). Where d_{ij} represents the distance measure from object i to cluster prototype j and d_{ik} represents the distance measure from object i to cluster prototype k.

$$u_{ij} = \left(\sum_{k=1}^{C} \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}} \right)^{-1}, \forall i, \forall j \quad (0.4)$$

3. Recalculate cluster prototypes c_j based on membership values u_{ij} and attribute values o_i according to (0.5).

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot o_i}{\sum_{i=1}^{N} u_{ij}^m}, \ \forall j \quad (0.5)$$

4. Return to step 2 and repeat as desired

Iterative clustering learning based on object-cluster similarity metric algorithm (OCIL)

While the aforementioned EM and FCM algorithms are able to cluster mixed data sets, their suitability for this type of data set is not optimal. That is, the distribution of the various attribute values is not considered by the distance metric. Hence, similarity information that is actually present in the categorical attributes is not considered (Cheung & Jia, 2013). Consequently, an optimal clustering algorithm for mixed data sets should consider the information embedded in the categorical attributes and combine this with the distance information based on numerical attributes.

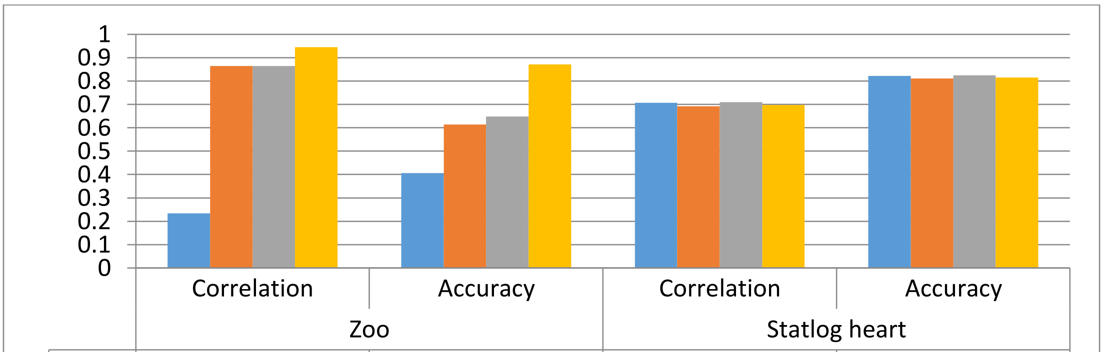
The OCIL algorithm, developed by Cheung and Jia (2013), is specifically aimed at clustering mixed-attribute data. The algorithm is non-fuzzy and similar to the k-means algorithm[8]. Likewise, when applied to purely numerical data sets, the OCIL algorithm is equivalent to the k-means algorithm (Cheung & Jia, 2013). In addition, the algorithm does not require any input parameters apart from the number of clusters, though this can also be calculated using the PLC-OC algorithm presented in the same paper (Cheung & Jia, 2013).

The OCIL algorithm employs a similarity metric specifically designed for mixed data sets. This metric consists of two separate metrics for the categorical and the numerical parts of the data. Based on its importance, a weight is assigned to each categorical attribute. In short, this means that a categorical attribute with many different values provides a lot of information and thus has a high weight. Conversely, a categorical attribute for which each object in the data set has the same value, is considered to be of zero importance. The similarity metric for the categorical attributes is combined with the similarity metric for numerical attributes (which is the Euclidean distance). This yields the object cluster similarity metric.

Due to the more complex nature of the algorithm as compared to the previous two algorithms, the OCIL algorithm is not presented here. Instead, the reader is referred to the original work of Cheung and Jia (2013) for a comprehensive overview of the algorithm.

Clustering results on test data

The aforementioned algorithms have been tested on two often used data sets in the field of machine learning, acquired from the UCI Machine Learning Repository (Lichman, 2013). Two data sets, Zoo and Statlog (heart), with mixed variable types have been used to resemble the type of data that is to be expected in behavioural classification in an airport situation. For a full description of the data set, the reader is referred UCI Machine Learning Repository website[9]. The results of the various algorithms on the test data is shown in the figure below.



	Zoo Correlation	Zoo Accuracy	Statlog heart Correlation	Statlog heart Accuracy
EM	0.2331	0.4059	0.7066	0.8222
FCM	0.8646	0.6139	0.6924	0.8111
OCIL	0.8646	0.6477	0.7093	0.8243
LCA	0.945	0.8713	0.6971	0.8148

[8] K-means is an algorithm for hard clustering, which is very similar to the fuzzy c-means algorithm in the limit of fuzziness parameter $m \rightarrow 1$

[9] https://archive.ics.uci.edu/ml/index.html

This page intentionally left blank.

Appendix C
Description of the CC-framework implementation

Performing the clustering and classification is implemented in the R programming language. The script that is used for this is described here. This description is roughly based on the sections in the used script source. Sections in the source can be recognized by the four preceding and four trailing hashes on a line.

1. The environment is initialized:
   a. Loading required packages:
      i. poLCA version 1.4.1
      ii. statar version 0.6.2
      iii. ggplot2 version 2.1.0
      iv. adabag version 4.1
      v. Hmisc version 3.17.4
      vi. pROC version 1.8
   b. Required functions are loaded:
      i. func_classification.R for the classification
      ii. func_binning.R for binning continuous variables
      iii. func_LCA.R for estimating the latent class model
2. Settings with respect to binning, LCA and classification are set. Additionally, the random seed is fixed if reproducible results are desired. The following parameters can be set:
   a. Binning
      i. The bin size in minutes
   b. LCA
      i. Number of classes/maximum number of classes
      ii. Whether or not the optimum number of classes should be found, based on BIC
      iii. The number of times the LCA should be estimated, in order to avoid a local optimum
      iv. Whether a CSV-file with the LCA results should be produced
      v. Whether text with the LCA results should be written to the screen
      vi. Whether plots with the LCA results should made
   c. Classification
      i. Whether a classifier should be created
      ii. The maximum number of decision trees to use in the classifier
      iii. The maximum depth of the decision trees in the classifier
      iv. The size of the training set (as a fraction of the total set)
      v. Whether text results should be written to the screen
      vi. Whether plots with results should be made
3. Data preparation
   a. Data is loaded from a CSV file
   b. Continuous attributes are binned using the func_binning.R function
4. Performing the LCA
   a. The attributes from the data to be used in the LCA are defined
   b. The LCA is ran using the func_LCA.R function
   c. Optionally, plots with the LCA results are made
   d. Optionally, a CSV with the original data, appended with the found classes is created
5. Performing the ensemble classification
   a. The attributes to be used in the classification are defined
   b. The classifier is built using the func_LCA.R function

i. The func_confmatrix.R function is loaded and executed from this function to obtain the performance metrics for the classifier

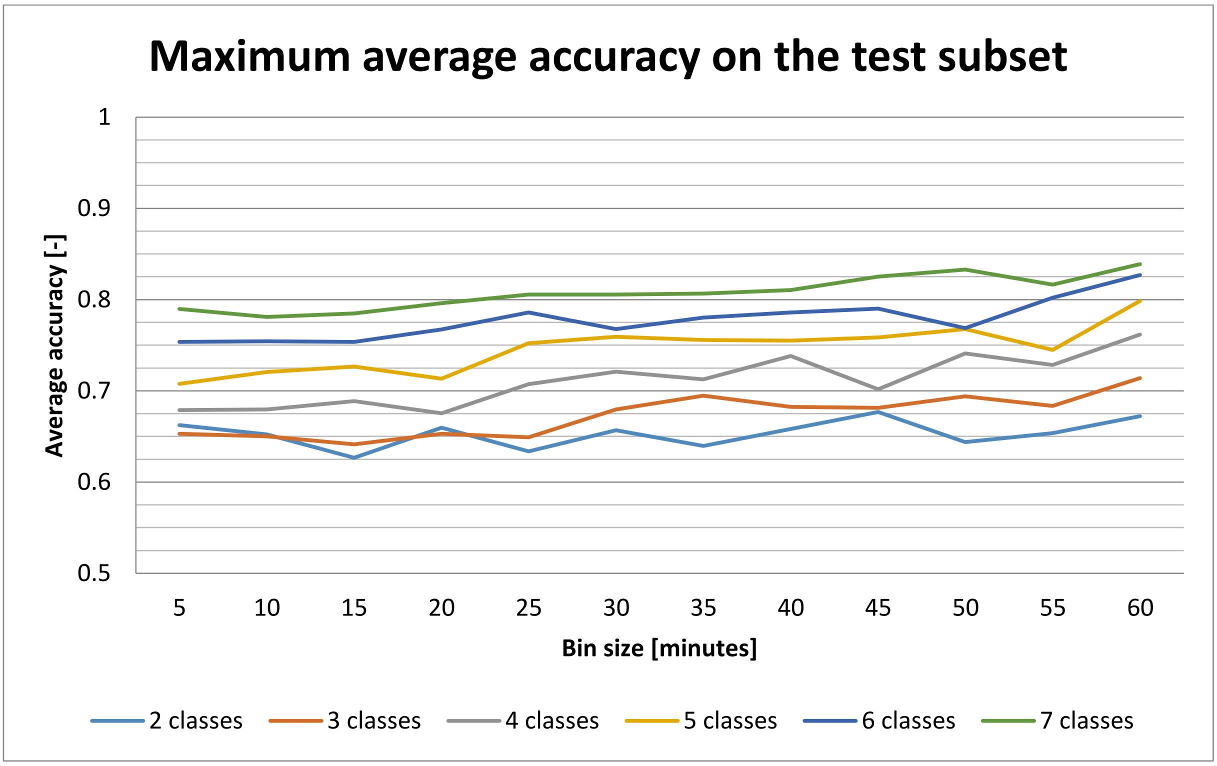
Appendix D
PASSME data set information

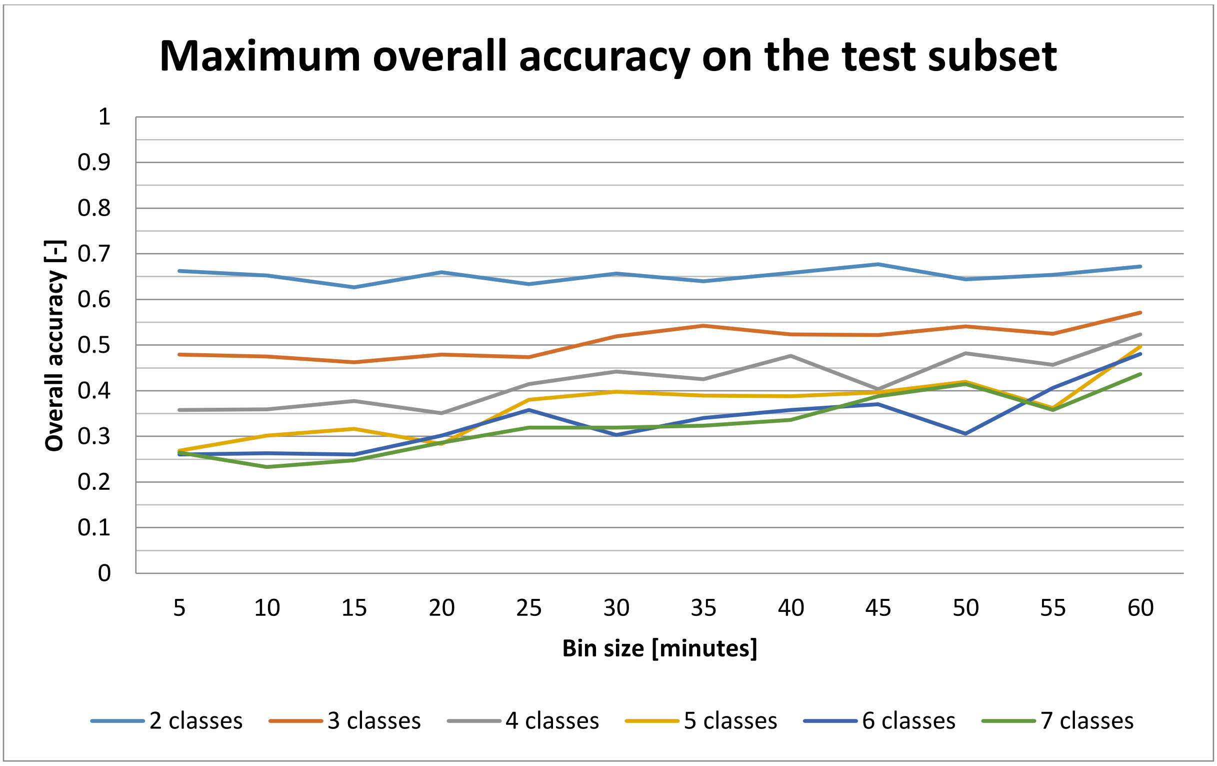
This appendix has been omitted due to the confidentiality of the data.

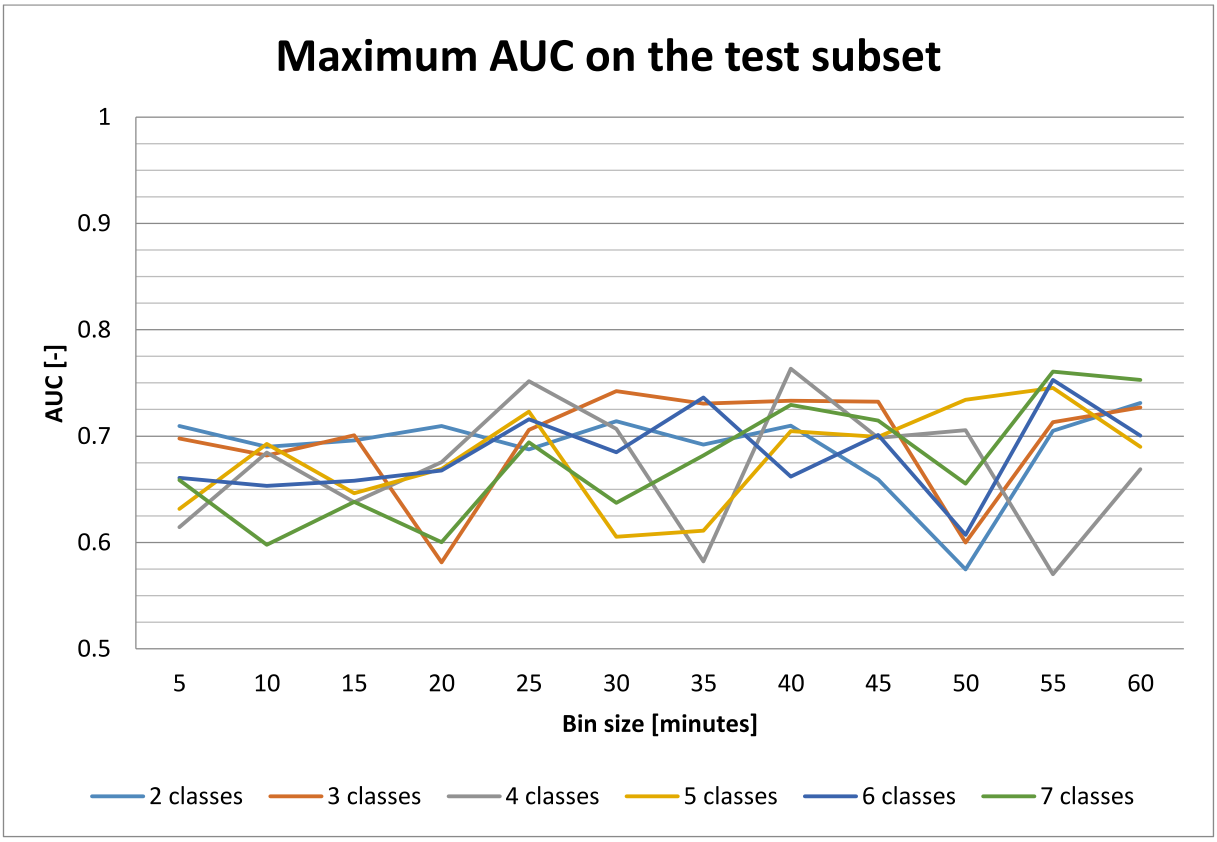
This page intentionally left blank.

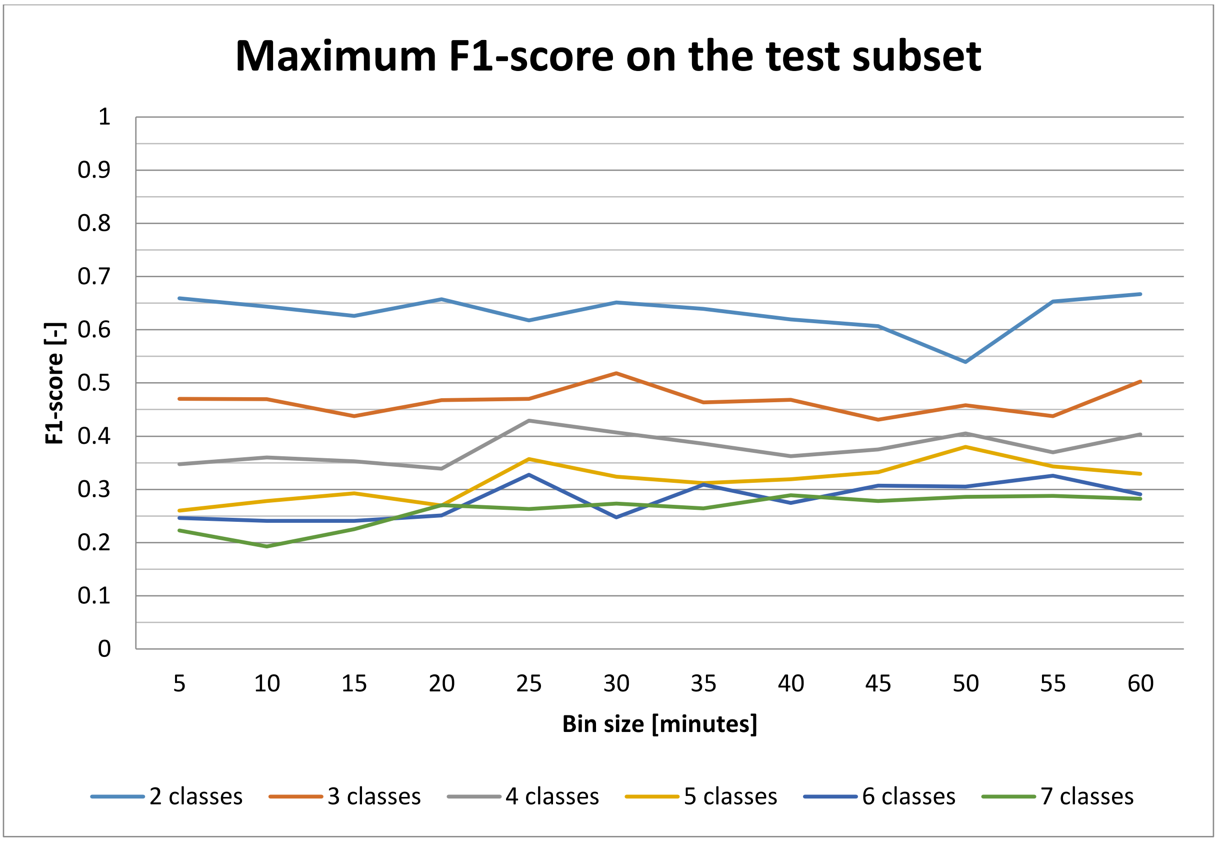
Appendix E
CC results on the PASSME data set

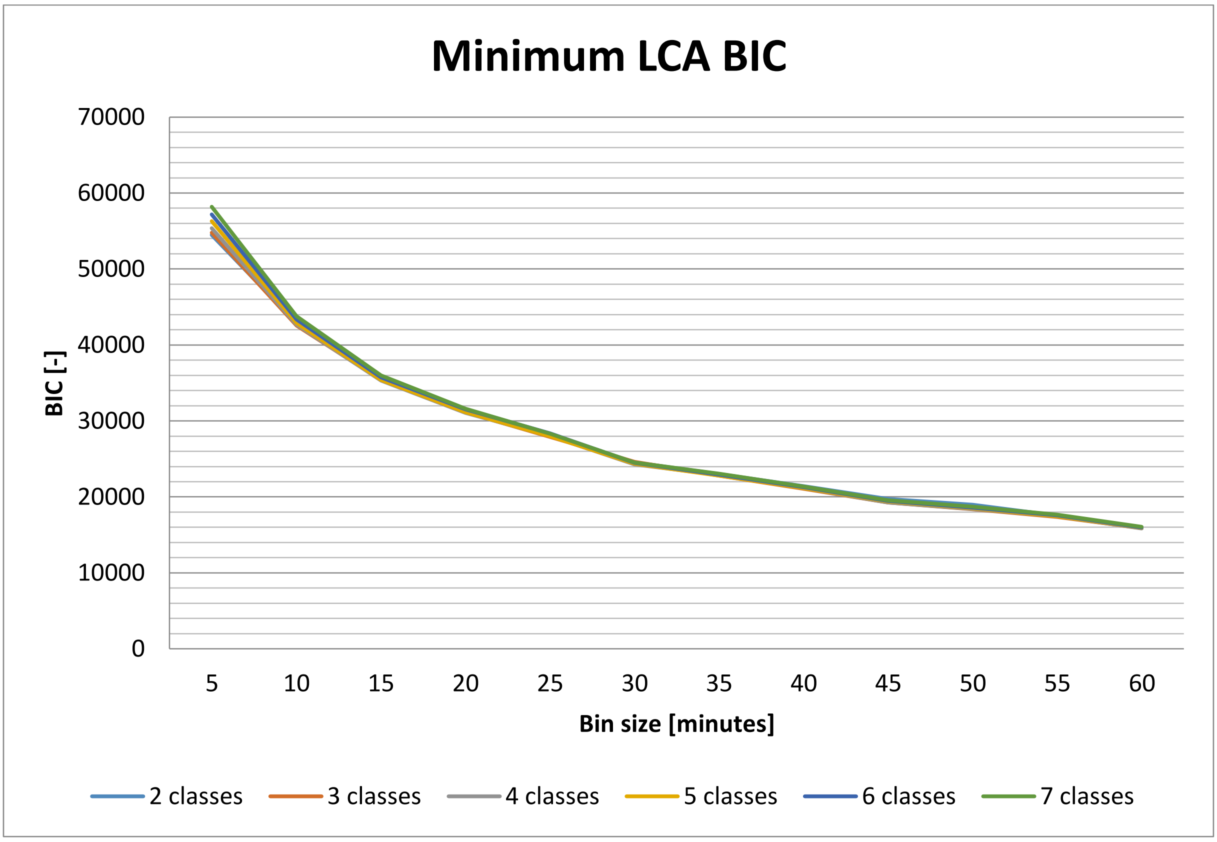
Results for the departing passengers subset

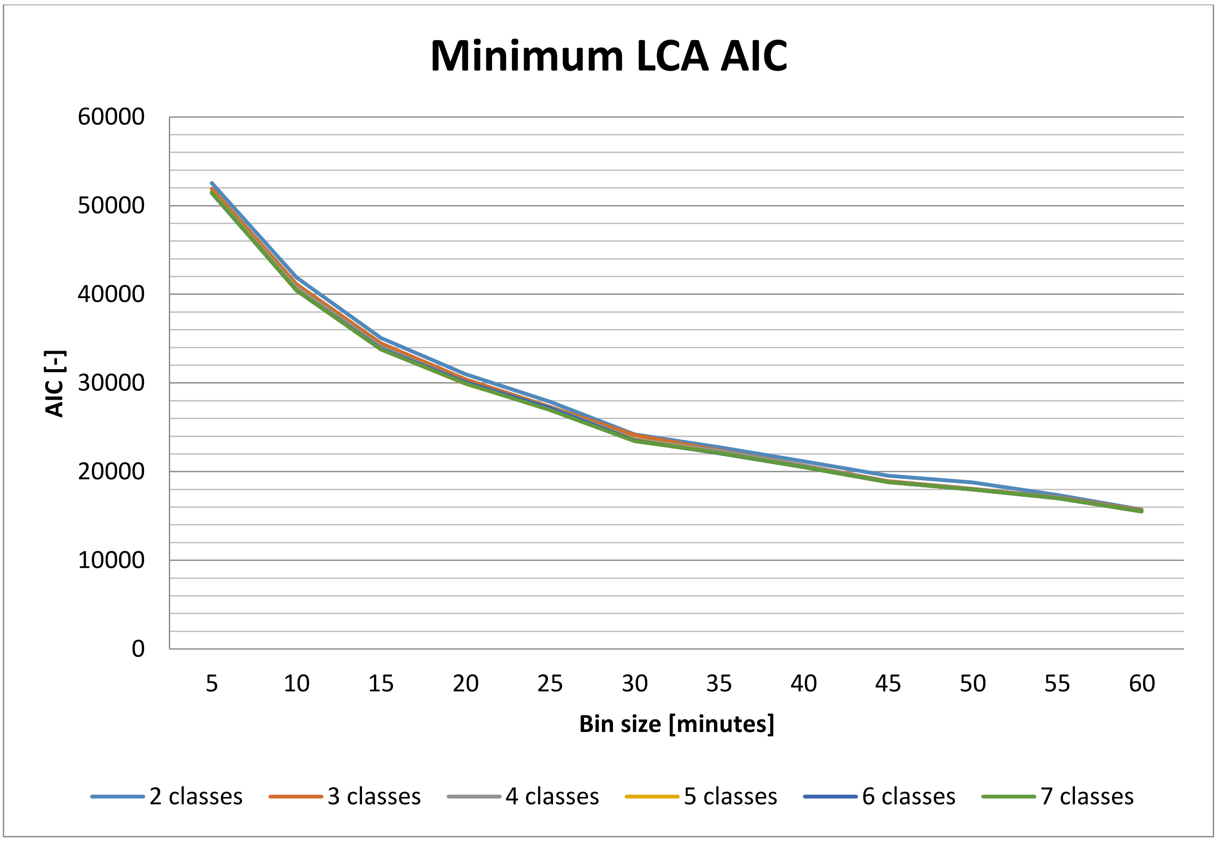


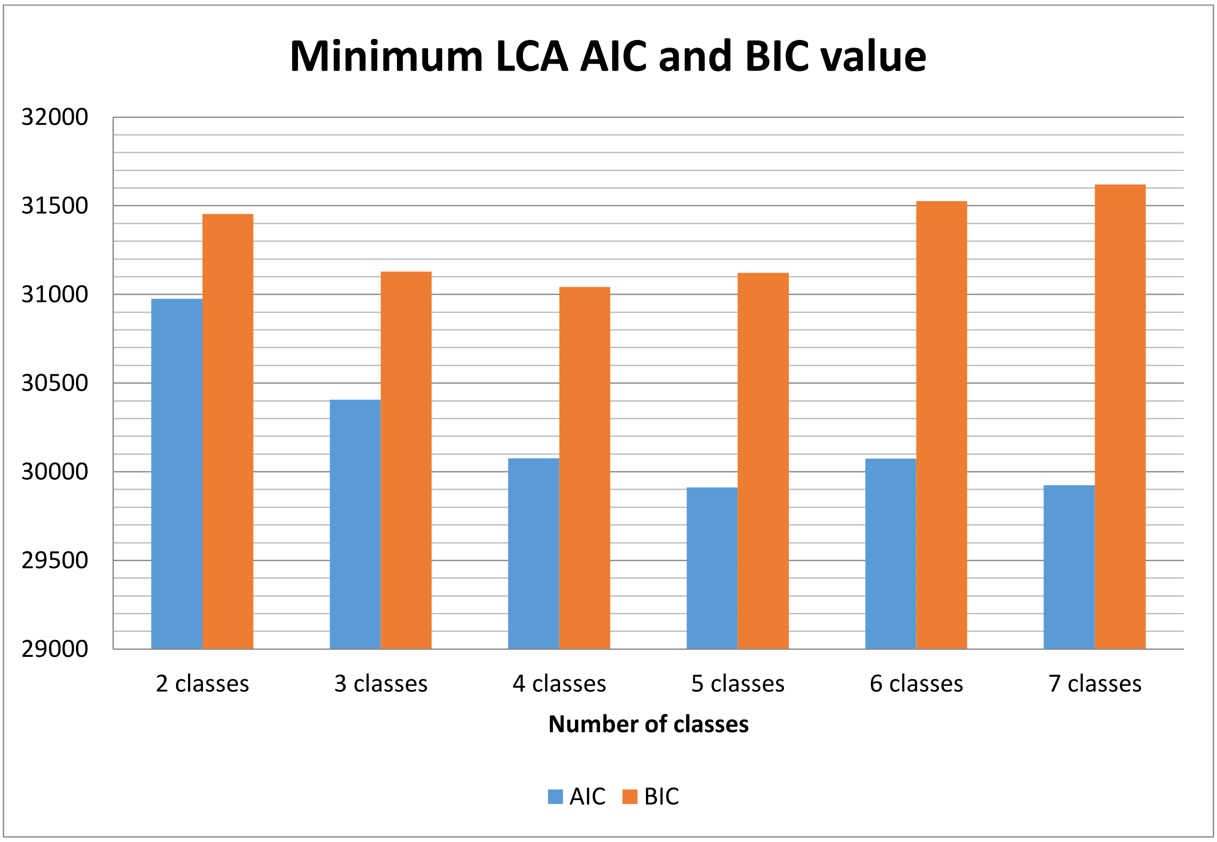


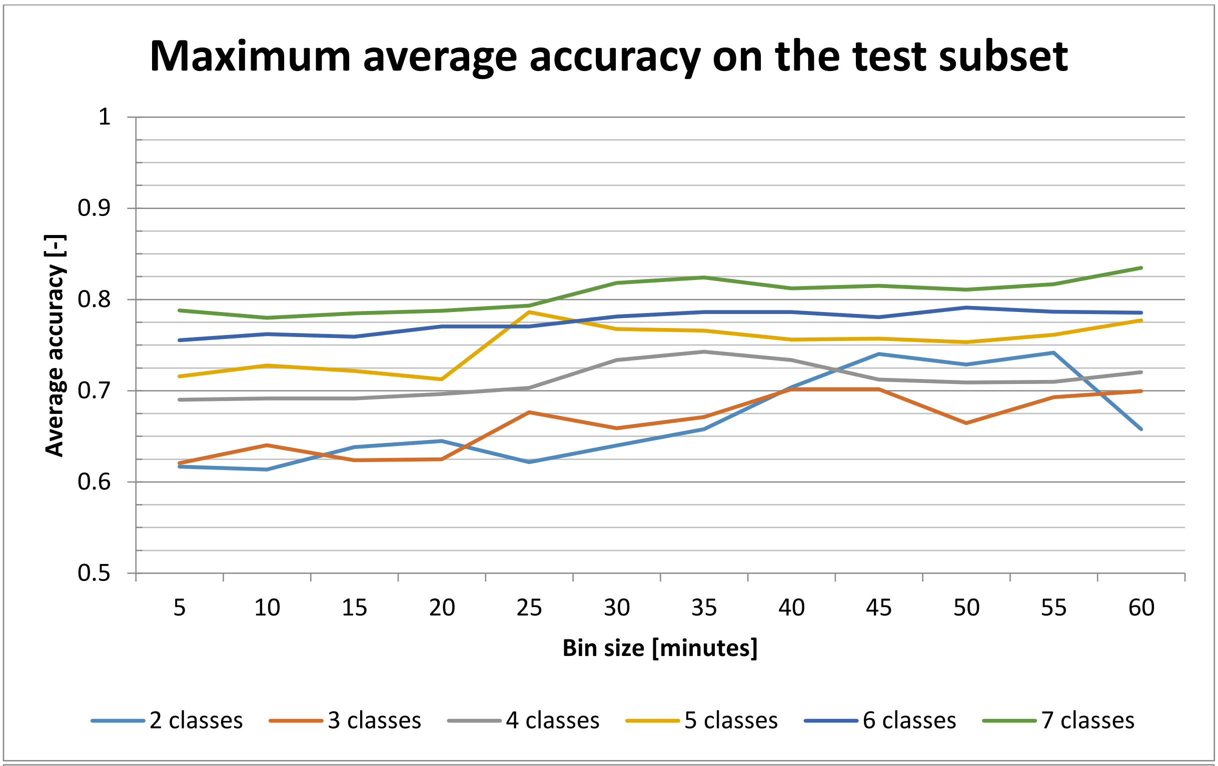

Maximum AUC on the test subset
1
0.9
0.8
0.7
0.6
0.5
AUC [-]
5
10
15
20
25
30
35
40
45
50
55
60
Bin size [minutes]
2 classes
3 classes
4 classes
5 classes
6 classes
7 classes

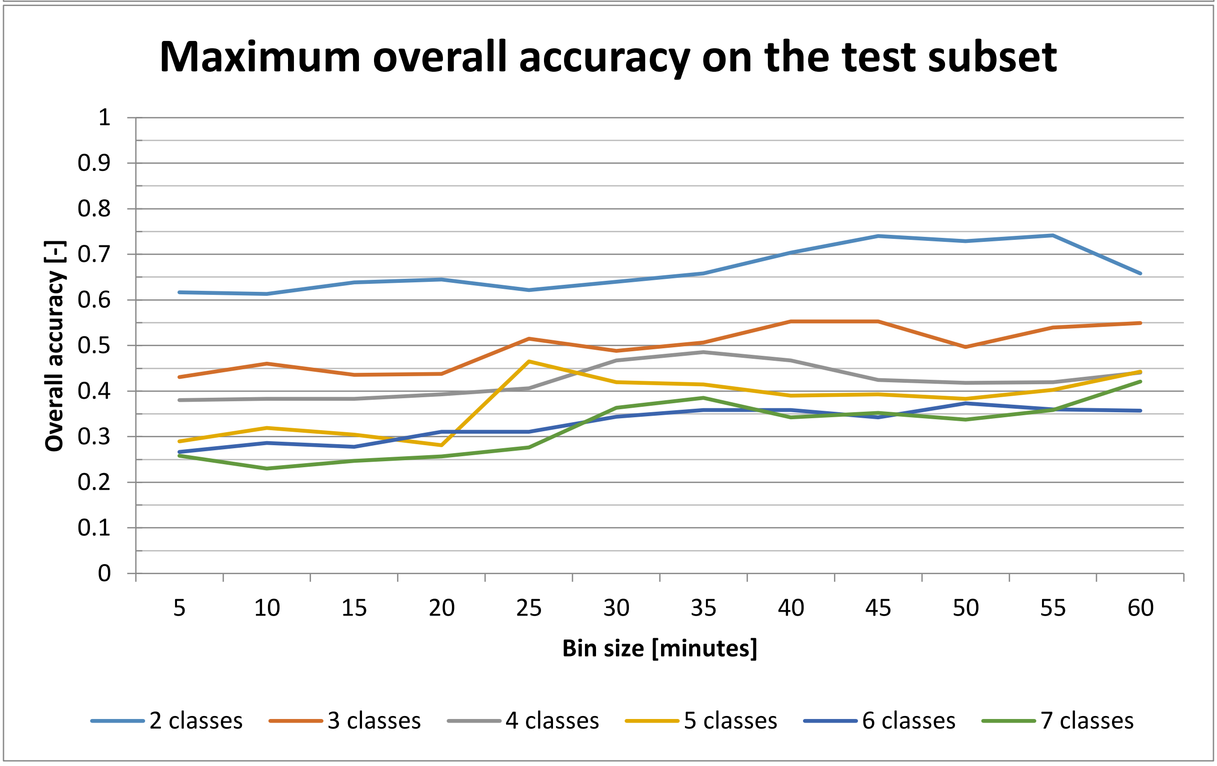

Maximum F1-score on the test subset
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
F1-score [-]
5
10
15
20
25
30
35
40
45
50
55
60
Bin size [minutes]
2 classes
3 classes
4 classes
5 classes
6 classes
7 classes

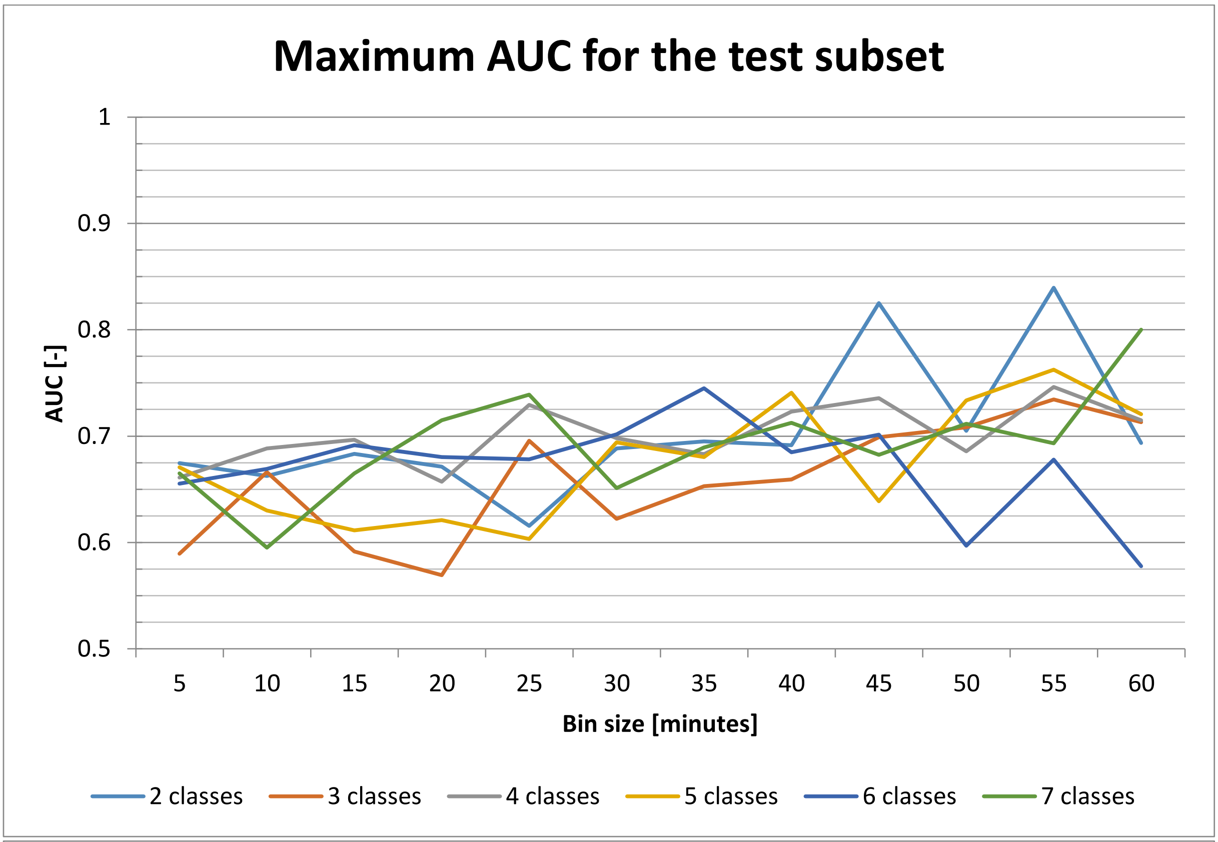

Minimum LCA BIC
70000
60000
50000
40000
30000
20000
10000
0
BIC [-]
5
10
15
20
25
30
35
40
45
50
55
60
Bin size [minutes]
2 classes
3 classes
4 classes
5 classes
6 classes
7 classes

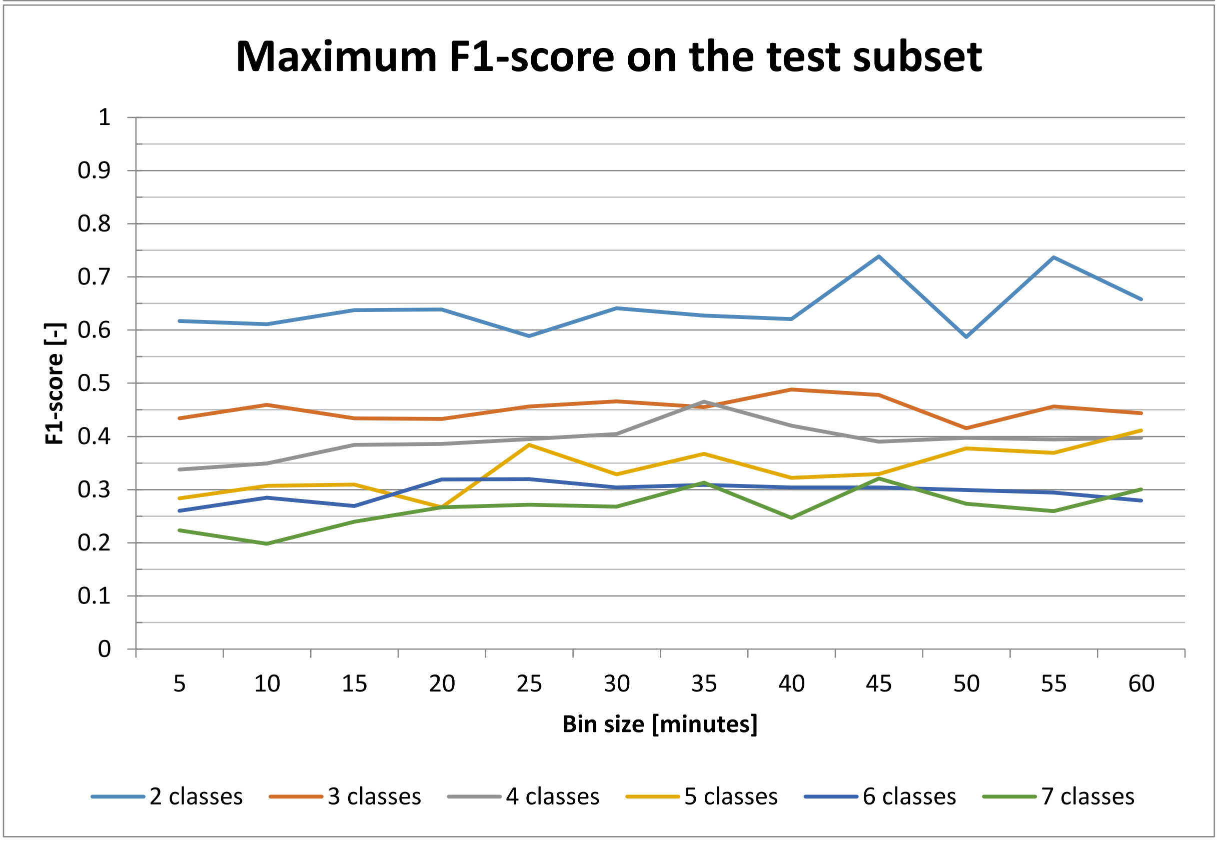

Minimum LCA AIC
60000
50000
40000
30000
20000
10000
0
AIC [-]
5
10
15
20
25
30
35
40
45
50
55
60
Bin size [minutes]
2 classes
3 classes
4 classes
5 classes
6 classes
7 classes

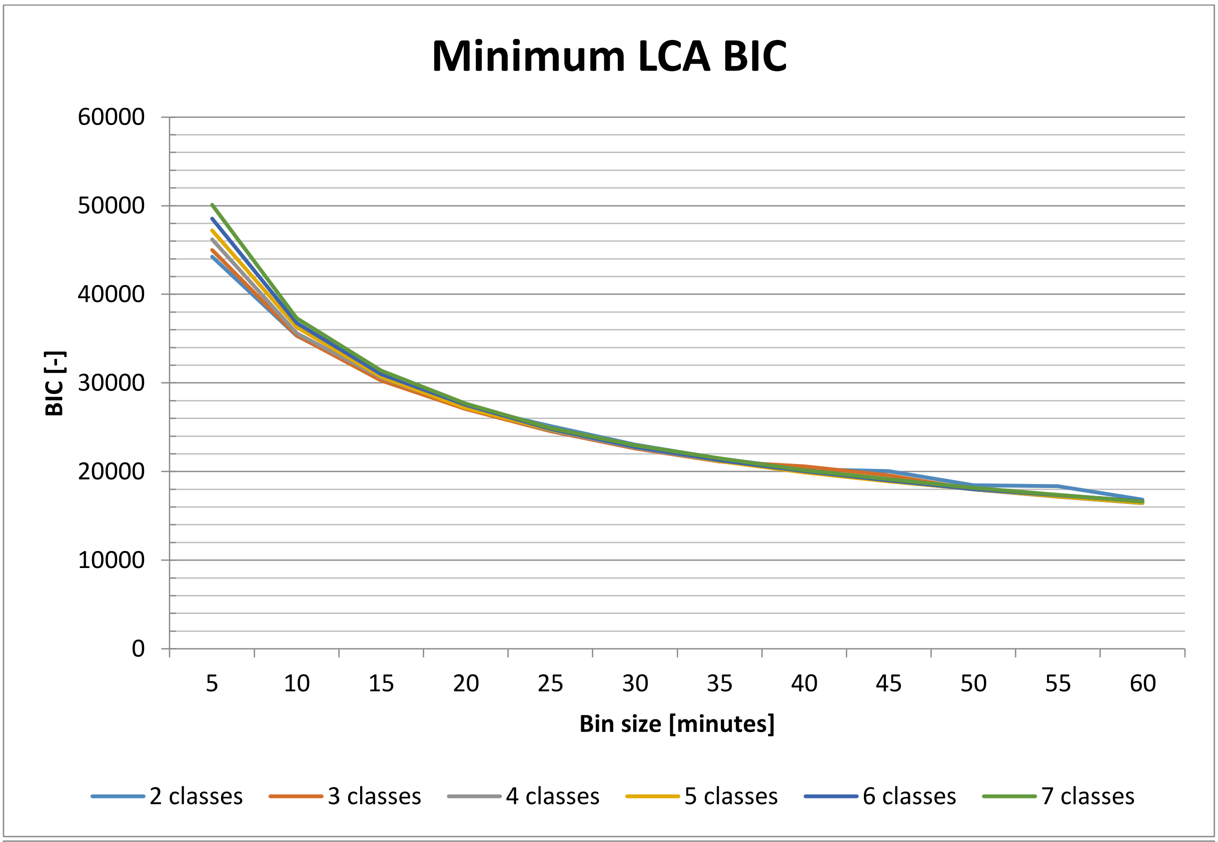

Minimum LCA AIC and BIC value
32000
31500
31000
30500
30000
29500
29000
2 classes
3 classes
4 classes
5 classes
6 classes
7 classes
Number of classes
AIC
BIC

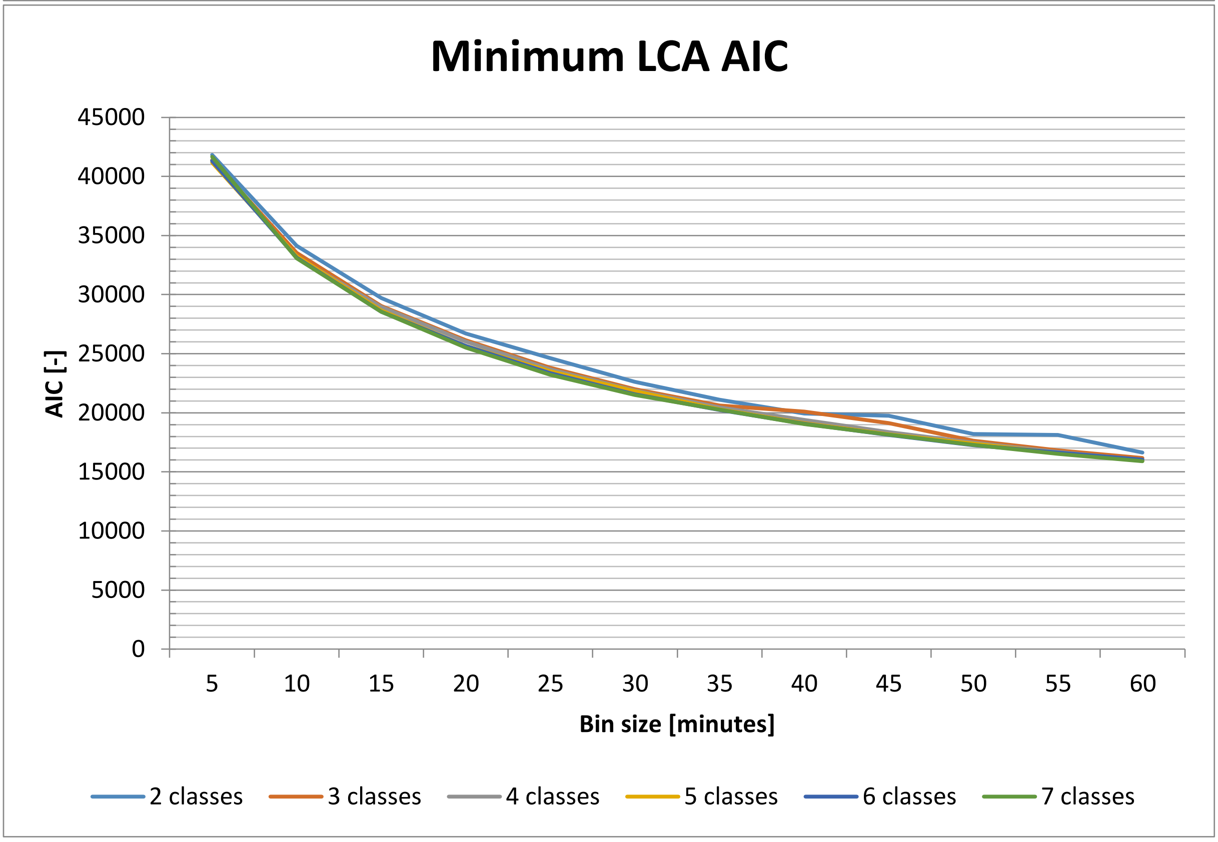

Maximum average accuracy on the test subset
1
0.9
0.8
0.7
0.6
0.5
Average accuracy [-]
5
10
15
20
25
30
35
40
45
50
55
60
Bin size [minutes]
2 classes
3 classes
4 classes
5 classes
6 classes
7 classes

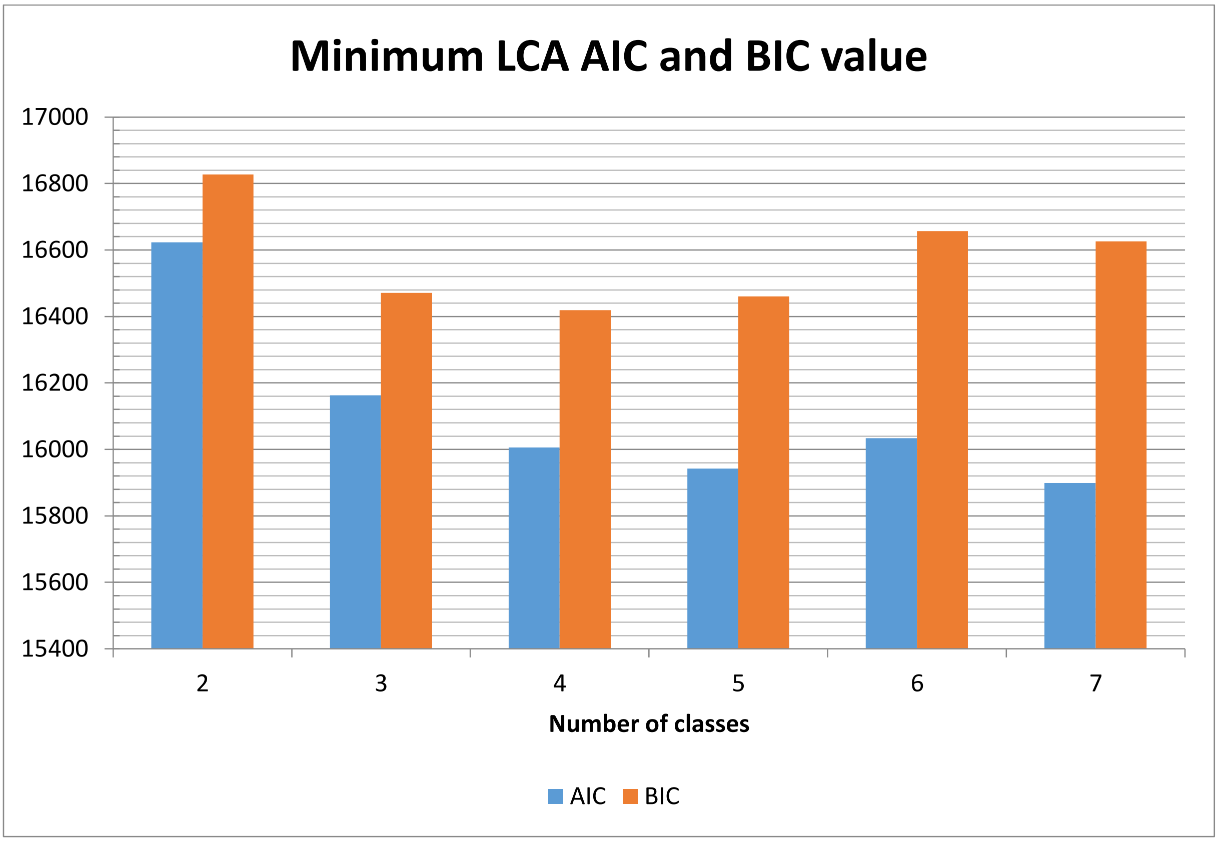

Maximum overall accuracy on the test subset
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Overall accuracy [-]
5
10
15
20
25
30
35
40
45
50
55
60
Bin size [minutes]
2 classes
3 classes
4 classes
5 classes
6 classes
7 classes


Maximum AUC for the test subset
AUC [-]
1
0.9
0.8
0.7
0.6
0.5
5
10
15
20
25
30
35
40
45
50
55
60
Bin size [minutes]
2 classes
3 classes
4 classes
5 classes
6 classes
7 classes


Maximum F1-score on the test subset
F1-score [-]
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
15
20
25
30
35
40
45
50
55
60
Bin size [minutes]
2 classes
3 classes
4 classes
5 classes
6 classes
7 classes


Minimum LCA BIC
60000
50000
40000
30000
20000
10000
0
BIC [-]
5
10
15
20
25
30
35
40
45
50
55
60
Bin size [minutes]
2 classes
3 classes
4 classes
5 classes
6 classes
7 classes


Minimum LCA AIC
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
AIC [-]
5
10
15
20
25
30
35
40
45
50
55
60
Bin size [minutes]
2 classes
3 classes
4 classes
5 classes
6 classes
7 classes


Minimum LCA AIC and BIC value
17000
16800
16600
16400
16200
16000
15800
15600
15400
2
3
4
5
6
7
Number of classes
AIC
BIC